(19)日本国特許庁（ＪＰ）　(12)特　許　公　報（Ｂ２）　(11)特許番号
特許第3529049号
(P3529049)

(45)発行日　平成16年５月24日(2004.5.24)　(24)登録日　平成16年３月５日(2004.3.5)

(51)Int.Cl.'	識別記号	FI	
Ｇ１０Ｌ　15/22		Ａ６３Ｈ　11/00	Ｚ
Ａ６３Ｈ　11/00		Ｂ２５Ｊ　5/00	Ｆ
Ｂ２５Ｊ　5/00		Ｇ１０Ｌ　3/00	５７１Ｕ
Ｇ１０Ｌ　13/00			５２１Ｊ
15/00			５２１Ｆ

請求項の数15(全 27 頁)　最終頁に続く

(21)出願番号　特願2002－60425(P2002－60425)

(22)出願日　平成14年３月６日(2002.3.6)

(65)公開番号　特開2003－255989(P2003－255989A)

(43)公開日　平成15年９月10日(2003.9.10)

審査請求日　平成15年９月12日(2003.9.12)

(73)特許権者　000002185
ソニー株式会社
東京都品川区北品川６丁目７番35号

(72)発明者　下村　秀樹
東京都品川区北品川６丁目７番35号ソニー株式会社内

(72)発明者　青山　一美
東京都品川区北品川６丁目７番35号ソニー株式会社内

(72)発明者　山田　敬一
東京都品川区北品川６丁目７番35号ソニー株式会社内

(74)代理人　100082740
弁理士　田辺　恵基

審査官　樫本　剛

最終頁に続く

(54)【発明の名称】　学習装置及び学習方法並びにロボット装置

1

(57)【特許請求の範囲】

【請求項１】対話を通して対象とする物体の名前を取得する対話手段と、

上記対象とする物体の複数の特徴のデータを検出し、当該検出結果及び既知の物体の対応する特徴のデータに基づいて、上記対象とする物体を認識する複数の認識手段と、

上記既知の物体の名前に対する各上記認識手段の認識結果を関連付けた関連付け情報を記憶する記憶手段と、

上記対話手段が取得した上記対象とする物体の名前、上記対象とする物体に対する各上記認識手段の認識結果、及び上記記憶手段が記憶する上記関連付け情報に基づいて、上記対象とする物体が新規な物体であるか否かを判断する判断手段と、

上記判断手段が上記対象とする物体を新規な物体と判断

2

したときに、当該対象とする物体に対応する上記複数の特徴のデータを各上記認識手段に記憶させると共に、当該対象とする物体についての関連付け情報を上記記憶手段に記憶させる制御手段とを具えることを特徴とする学習装置。

【請求項２】上記制御手段は、

上記判断手段が上記対象とする物体を上記既知の物体であると判断したときに、当該対象とする物体を正しく認識できた上記認識手段を、追加学習するよう制御することを特徴とする請求項１に記載の学習装置。

【請求項３】上記制御手段は、

上記判断手段が上記対象とする物体を上記既知の物体であると判断したときに、当該対象とする物体を正しく認識できなかった上記認識手段を、訂正学習するよう制御することを特徴とする請求項１に記載の学習装置。

【請求項４】上記判断手段は、
上記記憶手段が記憶する上記関連付け情報を参照しながら、上記対話手段が取得した上記対象とする物体の名前及び当該物体に対する各上記認識手段の認識結果の多数決により、上記対象とする物体が新規な物体であるか否かを判断することを特徴とする請求項１に記載の学習装置。
【請求項５】上記制御手段は、
必要に応じて対話を引き伸ばすように上記対話手段を制御することを特徴とする請求項１に記載の学習装置。
【請求項６】対話を通して対象とする物体の名前を取得する対話ステップと、
上記対象とする物体の複数の特徴のデータを検出し、当該検出結果及び既知の物体の対応する特徴のデータに基づいて、上記対象とする物体を認識する複数の認識ステップと、
上記既知の物体の名前に対する各上記認識手段の認識結果を関連付けた関連付け情報を記憶する記憶ステップと、
上記対話手段が取得した上記対象とする物体の名前、上記対象とする物体に対する各上記認識手段の認識結果、及び上記記憶手段が記憶する上記関連付け情報に基づいて、上記対象とする物体が新規な物体であるか否かを判断する判断ステップと、
上記判断手段が上記対象とする物体を新規な物体と判断したときに、当該対象とする物体に対応する上記複数の特徴のデータを各上記認識手段に記憶させると共に、当該対象とする物体についての関連付け情報を上記記憶手段に記憶させる制御ステップとを具えることを特徴とする学習方法。
【請求項７】上記制御ステップでは、
上記対象とする物体を上記既知の物体であると判断したときに、当該対象とする物体を正しく認識できた上記特徴について、追加学習することを特徴とする請求項６に記載の学習方法。
【請求項８】上記制御ステップでは、
上記対象とする物体を上記既知の物体であると判断したときに、当該対象とする物体を正しく認識できなかった上記特徴について、訂正学習することを特徴とする請求項６に記載の学習方法。
【請求項９】上記判断ステップでは、
上記関連付け情報を参照しながら、取得した上記対象とする物体の名前及び当該物体の各上記特徴にそれぞれ基づく各認識結果の多数決により、上記対象とする物体が新規な物体であるか否かを判断することを特徴とする請求項６に記載の学習方法。
【請求項１０】上記対話ステップでは、
必要に応じて当該対話を引き伸ばすことを特徴とする請求項６に記載の学習方法。
【請求項１１】対話を通して対象とする物体の名前を取得する対話手段と、
上記対象とする物体の複数の特徴のデータを検出し、当該検出結果及び既知の物体の対応する特徴のデータに基づいて、上記対象とする物体を認識する複数の認識手段と、
上記既知の物体の名前に対する各上記認識手段の認識結果を関連付けた関連付け情報を記憶する記憶手段と、
上記対話手段が取得した上記対象とする物体の名前、上記対象とする物体に対する各上記認識手段の認識結果、及び上記記憶手段が記憶する上記関連付け情報に基づいて、上記対象とする物体が新規な物体であるか否かを判断する判断手段と、
上記判断手段が上記対象とする物体を新規な物体と判断したときに、当該対象とする物体に対応する上記複数の特徴のデータを各上記認識手段に記憶させると共に、当該対象とする物体についての関連付け情報を上記記憶手段に記憶させる制御手段とを具えることを特徴とするロボット装置。
【請求項１２】上記制御手段は、
上記判断手段が上記対象とする物体を上記既知の物体であると判断したときに、当該対象とする物体を正しく認識できた上記認識手段を、追加学習するよう制御することを特徴とする請求項１１に記載のロボット装置。
【請求項１３】上記制御手段は、
上記判断手段が上記対象とする物体を上記既知の物体であると判断したときに、当該対象とする物体を正しく認識できなかった上記認識手段を、訂正学習するよう制御することを特徴とする請求項１１に記載のロボット装置。
【請求項１４】上記判断手段は、
上記記憶手段が記憶する上記関連付け情報を参照しながら、上記対話手段が取得した上記対象とする物体の名前及び当該物体に対する各上記認識手段の認識結果の多数決により、上記対象とする物体が上記新規な物体であるか否かを判断することを特徴とする請求項１１に記載のロボット装置。
【請求項１５】上記制御手段は、
必要に応じて対話を引き伸ばすように上記対話手段を制御することを特徴とする請求項１１に記載のロボット装置。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は学習装置及び学習方法並びにロボット装置に関し、例えばエンターテインメントロボットに適用して好適なものである。
【０００２】
【従来の技術】近年、一般家庭向けのエンターテインメントロボットが数多く商品化されている。そしてこのようなエンターテインメントロボットの中には、ＣＣＤ（ChargeCoupled Device）カメラやマイクロホン等の各

種外部センサが搭載され、これら外部センサの出力に基づいて外部状況を認識し、認識結果に基づいて自律的に行動し得るようになされたものなどもある。

【０００３】

【発明が解決しようとする課題】ところで、かかるエンターテインメントロボットにおいて、新規な物体（人物も含む。以下、同じ。）の名前をその物体と対応付けて覚えられるようにすることができれば、ユーザとのコミュニケーションをより円滑にすることができ、またユーザからの「ボールを蹴って」といった、予め名前が登録された物体以外の物体を対象とする種々の命令にも柔軟に対応し得るようにすることができるものと考えられる。なお、以下においては、上述のように物体の名前をその物体と対応付けて覚えることを『名前を学習する』と表現し、そのような機能を『名前学習機能』と呼ぶものとする。

【０００４】またこのような名前学習機能をエンターテインメントロボットに搭載するに際して、人間が普段行うように、エンターテインメントロボットが通常の人との対話を通して新規な物体の名前を学習できるようにすることができれば、その自然性から考えて最も望ましく、エンターテインメントロボットとしてのエンターテインメント性をより一層向上させ得るものと考えられる。

【０００５】ところが、従来技術では、名前を学習すべき新規の物体がいつ目の前に現れているのかをエンターテインメントロボットに判断させることが難しい問題がある。

【０００６】このため従来では、ユーザが明示的な音声コマンドを与え又はロボットに配設された特定のタッチセンサを押圧操作するなどして動作モードを登録モードに変更してから、物体の認識及びその名前の登録を行うといった手法が多く用いられている。しかしながら、ユーザとエンターテインメントロボットとの自然なインタラクションを考えるとき、このような明示的な指示による名前登録はいかにも不自然である問題があった。

【０００７】本発明は以上の点を考慮してなされたもので、エンターテイメント性を格段的に向上させ得る学習装置及び学習方法並びにロボット装置を提案しようとするものである。

【０００８】

【課題を解決するための手段】かかる課題を解決するため本発明においては、学習装置において、対話を通して対象とする物体の名前を取得する対話手段と、対象とする物体の複数の特徴のデータを検出し、当該検出結果及び既知の物体の対応する特徴のデータに基づいて、対象とする物体を認識する複数の認識手段と、既知の物体の名前に対する各認識手段の認識結果を関連付けた関連付け情報を記憶する記憶手段と、対話手段が取得した対象とする物体の名前、対象とする物体に対する各認識手段

の認識結果、及び記憶手段が記憶する関連付け情報に基づいて、対象とする物体が新規な物体であるか否かを判断する判断手段と、判断手段が対象とする物体を新規な物体と判断したときに、当該対象とする物体に対応する複数の特徴のデータを各認識手段に記憶させると共に、当該対象とする物体についての関連付け情報を記憶手段に記憶させる制御手段とを設けるようにした。

【０００９】この結果この学習装置は、音声コマンドの入力やタッチセンサの押圧操作等のユーザからの明示的な指示による名前登録を必要とすることなく、人間が普段行うように、通常の人との対話を通して新規な人物や物体等の名前を自然に学習することができる。

【００１０】また本発明においては、学習方法において、対話を通して対象とする物体の名前を取得する対話ステップと、対象とする物体の複数の特徴のデータを検出し、当該検出結果及び既知の物体の対応する特徴のデータに基づいて、対象とする物体を認識する複数の認識ステップと、既知の物体の名前に対する各認識手段の認識結果を関連付けた関連付け情報を記憶する記憶ステップと、対話手段が取得した対象とする物体の名前、対象とする物体に対する各認識手段の認識結果、及び記憶手段が記憶する関連付け情報に基づいて、対象とする物体が新規な物体であるか否かを判断する判断ステップと、判断手段が対象とする物体を新規な物体と判断したときに、当該対象とする物体に対応する複数の特徴のデータを各認識手段に記憶させると共に、当該対象とする物体についての関連付け情報を記憶手段に記憶させる制御ステップとを設けるようにした。

【００１１】この結果、この学習方法によれば、音声コマンドの入力やタッチセンサの押圧操作等のユーザからの明示的な指示による名前登録を必要とすることなく、人間が普段行うように、通常の人との対話を通して新規な人物や物体等の名前を自然に学習することができる。

【００１２】さらに本発明においては、ロボット装置において、対話を通して対象とする物体の名前を取得する対話手段と、対象とする物体の複数の特徴のデータを検出し、当該検出結果及び既知の物体の対応する特徴のデータに基づいて、対象とする物体を認識する複数の認識手段と、既知の物体の名前に対する各認識手段の認識結果を関連付けた関連付け情報を記憶する記憶手段と、対話手段が取得した対象とする物体の名前、対象とする物体に対する各認識手段の認識結果、及び記憶手段が記憶する関連付け情報に基づいて、対象とする物体が新規な物体であるか否かを判断する判断手段と、判断手段が対象とする物体を新規な物体と判断したときに、当該対象とする物体に対応する複数の特徴のデータを各認識手段に記憶させると共に、当該対象とする物体についての関連付け情報を記憶手段に記憶させる制御手段とを設けるようにした。

【００１３】この結果、このロボット装置は、音声コマ

ンドの入力やタッチセンサの押圧操作等のユーザからの明示的な指示による名前登録を必要とすることなく、人間が普段行うように、通常の人との対話を通して新規な人物や物体等の名前を自然に学習することができる。

【００１４】

【発明の実施の形態】以下図面について、本発明の一実施の形態を詳述する。

【００１５】（１）本実施の形態によるロボットの構成

図１及び図２において、１は全体として本実施の形態による２足歩行型のロボットを示し、胴体部ユニット２の上部に頭部ユニット３が配設されると共に、当該胴体部ユニット２の上部左右にそれぞれ同じ構成の腕部ユニット４Ａ、４Ｂがそれぞれ配設され、かつ胴体部ユニット２の下部左右にそれぞれ同じ構成の脚部ユニット５Ａ、５Ｂがそれぞれ所定位置に取り付けられることにより構成されている。

【００１６】胴体部ユニット２においては、体幹上部を形成するフレーム１０及び体幹下部を形成する腰ベース１１が腰関節機構１２を介して連結することにより構成されており、体幹下部の腰ベース１１に固定された腰関節機構１２の各アクチュエータA_1、A_2をそれぞれ駆動することによって、体幹上部を図３に示す直交するロール軸１３及びピッチ軸１４の回りにそれぞれ独立に回転させることができるようになされている。

【００１７】また頭部ユニット３は、フレーム１０の上端に固定された肩ベース１５の上面中央部に首関節機構１６を介して取り付けられており、当該首関節機構１６の各アクチュエータA_3、A_4をそれぞれ駆動することによって、図３に示す直交するピッチ軸１７及びヨー軸１８の回りにそれぞれ独立に回転させることができるようになされている。

【００１８】さらに各腕部ユニット４Ａ、４Ｂは、それぞれ肩関節機構１９を介して肩ベース１５の左右に取り付けられており、対応する肩関節機構１９の各アクチュエータA_5、A_6をそれぞれ駆動することによって図３に示す直交するピッチ軸２０及びロール軸２１の回りにそれぞれ独立に回転させることができるようになされている。

【００１９】この場合、各腕部ユニット４Ａ、４Ｂは、それぞれ上腕部を形成するアクチュエータA_7の出力軸に肘関節機構２２を介して前腕部を形成するアクチュエータA_8が連結され、当該前腕部の先端に手部２３が取り付けられることにより構成されている。

【００２０】そして各腕部ユニット４Ａ、４Ｂでは、アクチュエータA_7を駆動することによって前腕部を図３に示すヨー軸２４の回りに回転させ、アクチュエータA_8を駆動することによって前腕部を図３に示すピッチ軸２５の回りにそれぞれ回転させることができるようになされている。

【００２１】これに対して各脚部ユニット５Ａ、５Ｂにおいては、それぞれ股関節機構２６を介して体幹下部の腰ベース１１にそれぞれ取り付けられており、それぞれ対応する股関節機構２６の各アクチュエータをA_9～A_{11}それぞれ駆動することによって、図３に示す互いに直交するヨー軸２７、ロール軸２８及びピッチ軸２９の回りにそれぞれ独立に回転させることができるようになされている。

【００２２】この場合各脚部ユニット５Ａ、５Ｂは、それぞれ大腿部を形成するフレーム３０の下端に膝関節機構３１を介して下腿部を形成するフレーム３２が連結されると共に、当該フレーム３２の下端に足首関節機構３３を介して足部３４が連結されることにより構成されている。

【００２３】これにより各脚部ユニット５Ａ、５Ｂにおいては、膝関節機構３１を形成するアクチュエータA_{12}を駆動することによって、下腿部を図３に示すピッチ軸３５の回りに回転させることができ、また足首関節機構３３のアクチュエータA_{13}、A_{14}をそれぞれ駆動することによって、足部３４を図３に示す直交するピッチ軸３６及びロール軸３７の回りにそれぞれ独立に回転させることができるようになされている。

【００２４】一方、胴体部ユニット２の体幹下部を形成する腰ベース１１の背面側には、図４に示すように、当該ロボット１全体の動作制御を司るメイン制御部４０と、電源回路及び通信回路などの周辺回路４１と、バッテリ４５（図５）となどがボックスに収納されてなる制御ユニット４２が配設されている。

【００２５】そしてこの制御ユニット４２は、各構成ユニット（胴体部ユニット２、頭部ユニット３、各腕部ユニット４Ａ、４Ｂ及び各脚部ユニット５Ａ、５Ｂ）内にそれぞれ配設された各サブ制御部４３Ａ～４３Ｄと接続されており、これらサブ制御部４３Ａ～４３Ｄに対して必要な電源電圧を供給したり、これらサブ制御部４３Ａ～４３Ｄと通信を行ったりすることができるようになされている。

【００２６】また各サブ制御部４３Ａ～４３Ｄは、それぞれ対応する構成ユニット内の各アクチュエータA_1～A_{14}と接続されており、当該構成ユニット内の各アクチュエータA_1～A_{14}をメイン制御部４０から与えられる各種制御コマンドに基づいて指定された状態に駆動し得るようになされている。

【００２７】さらに頭部ユニット３には、図５に示すように、このロボット１の「目」として機能するＣＣＤ（Charge Coupled Device）カメラ５０及び「耳」として機能するマイクロホン５１及びタッチセンサ５２などからなる外部センサ部５３と、「口」として機能するスピーカ５４となどがそれぞれ所定位置に配設され、制御ユニット４２内には、バッテリセンサ５５及び加速度センサ５６などからなる内部センサ部５７が配設されている。

【００２８】そして外部センサ部５３のＣＣＤカメラ５０は、周囲の状況を撮像し、得られた画像信号Ｓ１Ａをメイン制御部に送出する一方、マイクロホン５１は、ユーザから音声入力として与えられる「歩け」、「伏せ」又は「ボールを追いかけろ」等の各種命令音声を集音し、かくして得られた音声信号Ｓ１Ｂをメイン制御部４０に送出するようになされている。

【００２９】またタッチセンサ５２は、図１及び図２において明らかなように頭部ユニット３の上部に設けられており、ユーザからの「撫でる」や「叩く」といった物理的な働きかけにより受けた圧力を検出し、検出結果を圧力検出信号Ｓ１Ｃとしてメイン制御部４０に送出する。

【００３０】さらに内部センサ部５７のバッテリセンサ５５は、バッテリ４５のエネルギ残量を所定周期で検出し、検出結果をバッテリ残量検出信号Ｓ２Ａとしてメイン制御部４０に送出する一方、加速度センサ５６は、３軸方向（ｘ軸、ｙ軸及びｚ軸）の加速度を所定周期で検出し、検出結果を加速度検出信号Ｓ２Ｂとしてメイン制御部４０に送出する。

【００３１】メイン制御部部４０は、外部センサ部５３のＣＣＤカメラ５０、マイクロホン５１及びタッチセンサ５２等からそれぞれ供給される画像信号Ｓ１Ａ、音声信号Ｓ１Ｂ及び圧力検出信号Ｓ１Ｃ等（以下、これらをまとめて外部センサ信号Ｓ１と呼ぶ）と、内部センサ部５７のバッテリセンサ５５及び加速度センサ等からそれぞれ供給されるバッテリ残量検出信号Ｓ２Ａ及び加速度検出信号Ｓ２Ｂ等（以下、これらをまとめて内部センサ信号Ｓ２と呼ぶ）に基づいて、ロボット１の周囲及び内部の状況や、ユーザからの指令、ユーザからの働きかけの有無などを判断する。

【００３２】そしてメイン制御部４０は、この判断結果と、予め内部メモリ４０Ａに格納されている制御プログラムと、そのとき装填されている外部メモリ５８に格納されている各種制御パラメータとに基づいて続く行動を決定し、決定結果に基づく制御コマンドを対応するサブ制御部４３Ａ～４３Ｄに送出する。この結果、この制御コマンドに基づき、そのサブ制御部４３Ａ～４３Ｄの制御のもとに、対応するアクチュエータA_1～A_{14}が駆動され、かくして頭部ユニット３を上下左右に揺動させたり、腕部ユニット４Ａ、４Ｂを上にあげたり、歩行するなどの行動がロボット１により発現されることとなる。

【００３３】またこの際メイン制御部４０は、必要に応じて所定の音声信号Ｓ３をスピーカ５４に与えることにより当該音声信号Ｓ３に基づく音声を外部に出力させたり、外見上の「目」として機能する頭部ユニット３の所定位置に設けられたＬＥＤに駆動信号を出力することによりこれを点滅させる。

【００３４】このようにしてこのロボット１において

は、周囲及び内部の状況や、ユーザからの指令及び働きかけの有無などに基づいて自律的に行動することができるようになされている。

【００３５】（２）名前学習機能に関するメイン制御部４０の処理

次にこのロボット１に搭載された名前学習機能について説明する。

【００３６】このロボット１には、人との対話を通してその人の名前を取得し、当該名前を、マイクロホン５１やＣＣＤカメラ５０の出力に基づいて検出したその人の声の音響的特徴及び顔の形態的特徴の各データと関連付けて記憶すると共に、これら記憶した各データに基づいて、名前を取得していない新規な人の登場を認識し、その新規な人の名前や声の音響的特徴及び顔の形態的特徴を上述と同様にして取得し記憶するようにして、人の名前をその人と対応付けて取得（以下、これを名前の学習と呼ぶ）学習していく名前学習機能が搭載されている。なお以下においては、その人の声の音響的特徴及び顔の形態的特徴と対応付けて名前を記憶し終えた人を『既知の人』と呼び、記憶し終えていない人を『新規な人』と呼ぶものとする。

【００３７】そしてこの名前学習機能は、メイン制御部４０における各種処理により実現されている。

【００３８】ここで、かかる名前学習機能に関するメイン制御部４０の処理内容を機能的に分類すると、図６に示すように、人が発声した言葉を認識する音声認識部６０と、人の声の音響的特徴を検出すると共に当該検出した音響的特徴に基づいてその人を識別して認識する話者認識部６１と、人の顔の形態的特徴を検出すると共に当該検出した形態的特徴に基づいてその人を識別して認識する顔認識部６２と、人との対話制御を含む新規な人の名前学習のための各種制御や、既知の人の名前、声の音響的特徴及び顔の形態的特徴の記憶管理を司る対話制御部６３と、対話制御部６３の制御のもとに各種対話用の音声信号Ｓ３を生成してスピーカ５４（図５）に送出する音声合成部６４とに分けることができる。

【００３９】この場合、音声認識部６０においては、マイクロホン５１（図５）からの音声信号Ｓ１Ｂに基づき所定の音声認識処理を実行することにより当該音声信号Ｓ１Ｂに含まれる言葉を単語単位で認識する機能を有するものであり、認識したこれら単語を文字列データＤ１として対話制御部６３に送出するようになされている。

【００４０】また話者認識部６１は、マイクロホン５１から与えられる音声信号Ｓ１Ｂに含まれる人の声の音響的特徴を、例えば“Segregation of Speakers for Recognition and Speaker Identification (CH2977-7/91/0000-0873 S1.00 1991 IEEE)”に記載された方法等を利用した所定の信号処理により検出する機能を有している。

【００４１】そして話者認識部６１は、通常時には、こ

の検出した音響的特徴のデータをそのとき記憶している全ての既知の人の音響的特徴のデータと順次比較し、そのとき検出した音響的特徴がいずれか既知の人の音響的特徴と一致した場合には当該既知の人の音響的特徴と対応付けられた当該音響的特徴に固有の識別子（以下、これをＳＩＤと呼ぶ）を対話制御部６３に通知する一方、検出した音響的特徴がいずれの既知の人の音響的特徴とも一致しなかった場合には、認識不能を意味するＳＩＤ（=－１）を対話制御部６３に通知するようになされている。

【００４２】また話者認識部６１は、対話制御部６３が新規な人であると判断したときに当該対話制御部６３から与えられる新規学習の開始命令及び学習終了命令に基づいて、その間その人の声の音響的特徴を検出し、当該検出した音響的特徴のデータを新たな固有のＳＩＤと対応付けて記憶すると共に、このＳＩＤを対話制御部６３に通知するようになされている。

【００４３】なお話者認識部６１は、対話制御部６３からの追加学習や訂正学習の開始命令及び終了命令に応じて、その人の声の音響的特徴のデータを追加的に収集する追加学習や、その人の声の音響的特徴のデータをその人を正しく認識できるよう訂正する訂正学習をも行い得るようになされている。

【００４４】顔認識部６２においては、ＣＣＤカメラ５０（図５）から与えられる画像信号Ｓ１Ａを常時監視し、当該画像信号Ｓ１Ａに基づく画像内に含まれる人の顔の形態的特徴を所定の信号処理により検出する機能を有している。

【００４５】そして顔認識部６２は、通常時には、この検出した形態的特徴のデータをそのとき記憶している全ての既知の人の顔の形態的特徴のデータと順次比較し、そのとき検出した形態的特徴がいずれか既知の人の顔の形態的特徴と一致した場合には当該既知の人の形態的特徴と対応付けられた当該形態的特徴に固有の識別子（以下、これをＦＩＤと呼ぶ）を対話制御部に通知する一方、検出した形態的特徴がいずれの既知の人の顔の形態的特徴とも一致しなかった場合には、認識不能を意味するＦＩＤ（=－１）を対話制御部に通知するようになされている。

【００４６】また顔認識部６２は、対話制御部６３が新規な人であると判断したときに当該対話制御部６３から与えられる学習開始命令及び学習終了命令に基づいて、その間ＣＣＤカメラ５０からの画像信号Ｓ１Ａに基づく画像内に含まれる人の顔の形態的特徴を検出し、当該検出した形態的特徴のデータを新たな固有のＦＩＤと対応付けて記憶すると共に、このＦＩＤを対話制御部６３に通知するようになされている。

【００４７】なお顔認識部６２は、対話制御部６３からの追加学習や訂正学習の開始命令及び終了命令に応じて、人の顔の形態的特徴のデータを追加的に収集する追加学習や、人の顔の形態的特徴のデータをその人を正しく認識できるよう訂正する訂正学習をも行い得るようになされている。

【００４８】音声合成部６４は、対話制御部６３から与えられる文字列データＤ２を音声信号Ｓ３に変換する機能を有し、かくして得られた音声信号Ｓ３をスピーカ５４（図５）に送出するようになされている。これによりこの音声信号Ｓ３に基づく音声をスピーカ５４から出力させることができるようになされている。

【００４９】対話制御部６３においては、図７に示すように、既知の人の名前と、話者認識部６１が記憶しているその人の声の音響的特徴のデータに対応付けられたＳＩＤと、顔認識部６２が記憶しているその人の顔の形態的特徴のデータに対応付けられたＦＩＤとを関連付けて記憶するメモリ６５（図６）を有している。

【００５０】そして対話制御部６３は、所定のタイミングで所定の文字列データＤ２を音声合成部６４に与えることにより、話し相手の人に対して名前を質問し又は名前を確認するための音声等をスピーカ５４から出力させる一方、このときのその人の応答等に基づく音声認識部６０及び話者認識部６１の各認識結果並びにその人に対する顔認識部６２の認識結果と、メモリ６５に格納された上述の既知の人の名前、ＳＩＤ及びＦＩＤの関連付けの情報とに基づいてその人が新規な人であるか否かを判断するようになされている。

【００５１】そして対話制御部６３は、その人が新規な人であると判断したときには、話者認識部６１及び顔認識部６２に対して新規学習の開始命令及び終了命令を与えることにより、これら話者認識部６１及び顔認識部６２にその新規な人の声の音響的特徴や顔の形態的特徴のデータを収集及び記憶させると共に、この結果としてこれら話者認識部６１及び顔認識部６２からそれぞれ与えられるその新規な人の声の音響的特徴のデータや顔の形態的特徴のデータに対応付けられたＳＩＤ及びＦＩＤを、かかる対話により得られたその人の名前と関連付けてメモリ６５に格納するようになされている。

【００５２】また対話制御部６３は、その人が既知の人であると判断したときには、必要に応じて話者認識部６１及び顔認識部６２に追加学習や訂正学習の開始命令を与えることにより話者認識部６１及び顔認識部６２に追加学習や訂正学習を行わせる一方、これと共に音声合成部６４に所定の文字列データＤ２を所定のタイミングで順次送出することにより、話者認識部６１及び顔認識部６２が追加学習や訂正学習をするのに必要な相当量のデータを収集できるまでその人との対話を長引かせるような対話制御を行うようになされている。

【００５３】（３）名前学習機能に関する対話制御部６３の具体的処理

次に、名前学習機能に関する対話制御部６３の具体的な処理内容について説明する。

【００５４】対話制御部６３は、外部メモリ５８（図５）に格納された制御プログラムに基づいて、図８及び図９に示す名前学習処理手順ＲＴ１に従って新規な人の名前を順次学習するための各種処理を実行する。

【００５５】すなわち対話制御部６３は、ＣＣＤカメラ５０からの画像信号Ｓ１Ａに基づき顔認識部６２が人の顔を認識することにより当該顔認識部６２からＦＩＤが与えられると名前学習処理手順ＲＴ１をステップＳＰ０において開始し、続くステップＳＰ１において、メモリ６５に格納された既知の人の名前と、これに対応するＳＩＤ及びこれに対応するＦＩＤとを関連付けた情報（以下、これを関連付け情報と呼ぶ）に基づいてそのＦＩＤから対応する名前を検索できるか否か（すなわちＦＩＤが認識不能を意味する「－１」でないか否か）を判断する。

【００５６】ここでこのステップＳＰ１において肯定結果を得ることは、その人が、顔認識部６２がその人の顔の形態的特徴のデータを記憶しており、当該データと対応付けられたＦＩＤがその人の名前と関連付けてメモリ６５に格納されている既知の人であることを意味する。ただしこの場合においても、顔認識部６２が新規の人を既知の人と誤認識したことも考えられる。

【００５７】そこで対話制御部６３は、ステップＳＰ１において肯定結果を得た場合には、ステップＳＰ２に進んで所定の文字列データＤ２を音声合成部６４に送出することにより、例えば図１０に示すように、「○○さんですよね。」といったその人の名前がＦＩＤから検索された名前（上述の○○に当てはまる名前）と一致するか否かを確かめるための質問の音声をスピーカ５４から出力させる。

【００５８】次いで対話制御部６３は、ステップＳＰ３に進んで、かかる質問に対するその人の「はい、そうです。」や「いいえ、違います。」といった応答の音声認識結果が音声認識部６０から与えられるのを待ち受ける。そして対話制御部６３は、やがて音声認識部６３からかかる音声認識結果が与えられ、また話者認識部６１からそのときの話者認識結果であるＳＩＤが与えられると、ステップＳＰ４に進んで、音声認識部６３からの音声認識結果に基づき、その人の応答が肯定的なものであるか否かを判断する。

【００５９】ここでこのステップＳＰ４において肯定結果を得ることは、ステップＳＰ１において顔認識部６２から与えられたＦＩＤに基づき検索された名前がその人の名前と一致しており、従ってその人は対話制御部６３が検索した名前を有する本人であるとほぼ断定できる状態にあることを意味する。

【００６０】かくしてこのとき対話制御部６３は、その人は当該対話制御部６３が検索した名前を有する本人であると断定し、ステップＳＰ５に進んで話者認識部６１に対して追加学習の開始命令を与える。またこれと共に対話制御部６３は、最初に話者認識部６１から与えられたＳＩＤが、かかる名前からメモリ６５に格納された関連付け情報に基づいて検索できるＳＩＤと一致している場合には話者認識部６１に対して追加学習の開始命令を与え、これに対して一致していない場合には訂正学習の開始命令を与える。

【００６１】そして対話制御部６３は、この後ステップＳＰ６に進んで例えば図１０のように「今日はいい天気ですね。」などといった、その人との対話を長引かせるための雑談をさせるための文字列データＤ２を音声合成部６４に順次送出し、この後追加学習又は訂正学習に十分な所定時間が経過すると、ステップＳＰ７に進んで話者認識部６１及び顔認識部６２に対して追加学習又は訂正学習の終了命令を与えた後、ステップＳＰ２０に進んでその人に対する名前学習処理を終了する。

【００６２】一方、ステップＳＰ１において否定結果を得ることは、顔認識部６２により顔認識された人が新規の人であるか、又は顔認識部６２が既知の人を新規の人と誤認識したことを意味する。またステップＳＰ４において否定結果を得ることは、最初に顔認識部６２から与えられたＦＩＤから検索された名前がその人の名前と一致していないことを意味する。そして、これらいずれの場合においても、対話制御部６３がその人を正しく把握していない状態にあるといえる。

【００６３】そこで対話制御部６３は、ステップＳＰ１において否定結果を得たときや、ステップＳＰ４において否定結果を得たときには、ステップＳＰ８に進んで音声合成部６４に文字列データＤ２を与えることにより、例えば図１１に示すように、「あれ、名前を教えてください。」といった、その人の名前を聞き出すための質問の音声をスピーカ５４から出力させる。

【００６４】そして対話制御部６３は、この後ステップＳＰ９に進んで、かかる質問に対するその人の「○○です。」といった応答の音声認識結果（すなわち名前）と、当該応答時における話者認識部６１の話者認識結果（すなわちＳＩＤ）とがそれぞれ音声認識部６０及び話者認識部６１から与えられるのを待ち受ける。

【００６５】そして対話制御部６３は、やがて音声認識部６０から音声認識結果が与えられ、話者認識部６１からＳＩＤが与えられると、ステップＳＰ１０に進んで、これら音声認識結果及びＳＩＤ並びに最初に顔認識部６２から与えられたＦＩＤに基づいて、その人が新規な人であるか否かを判断する。

【００６６】ここでこの実施の形態の場合、かかる判断は、音声認識部６０の音声認識により得られた名前と、話者認識部６１からのＳＩＤと、顔認識部６２からのＦＩＤとでなる３つの認識結果の多数決により行われる。

【００６７】例えば、話者認識部６１からのＳＩＤ及び顔認識部６２からのＦＩＤが共に認識不能を意味する「－１」で、かつステップＳＰにおいて音声認識部６０

からの音声認識結果に基づき得られたその人の名前がメモリ６５においてどのＳＩＤやＦＩＤとも関連付けられていない場合には、その人が新規な人であると判断する。既知のどの顔又はどの声とも似つかない人が全く新しい名前をもっているという状況であるので、そのような判断ができる。

【００６８】また対話制御部６３は、話者認識部６１からのＳＩＤ及び顔認識部６２からのＦＩＤがメモリ６５において異なる名前と関連付けられているか又はその一方が認識不能を意味する「－１」であり、かつステップＳＰ９において音声認識部６０からの音声認識結果に基づき得られたその人の名前がメモリ６５に格納されてない場合にも、その人が新規な人であると判断する。これは、各種認識処理において、新規カテゴリを既知カテゴリのどれかと誤認識するのは起こり易いことであり、また音声認識された名前が登録されていないことを考えれば、かなり高い確信度をもって新規の人と判断できるからである。

【００６９】これに対して対話制御部６３は、話者認識部６１からのＳＩＤ及び顔認識部６２からのＦＩＤがメモリ６５において同じ名前と関連付けられており、かつステップＳＰ９において音声認識部６０からの音声認識結果に基づき得られたその人の名前がそのＳＩＤ及びＦＩＤが関連付けられた名前である場合には、その人が既知の人であると判断する。

【００７０】また対話制御部６３は、話者認識部６１からのＳＩＤ及び顔認識部６２からのＦＩＤがメモリ６５において異なる名前と関連付けられており、かつステップＳＰ９において音声認識部６０からの音声認識結果に基づき得られたその人の名前がかかるＳＩＤ又はＦＩＤの一方が関連付けられた名前である場合には、その人が既知の人であると判断する。この場合は、話者認識部６１及び顔認識部６２のいずれか一方の認識結果が間違っていると考えられるため、かかる多数決によりそのように判断する。

【００７１】一方、対話制御部６３は、話者認識部６１からのＳＩＤ及び顔認識部６２からのＦＩＤがメモリ６５において異なる名前と関連付けられており、かつステップＳＰ９において音声認識部６０からの音声認識結果に基づき得られたその人の名前がメモリ６５においてかかるＳＩＤ及びＦＩＤのいずれにも関連付けられていない名前である場合には、その人が既知の人であるか又は新規の人であるかを判断しない。このケースでは、音声認識部６０、話者認識部６１及び顔認識部６２のいずれか又は全部の認識が間違っていることも考えられるが、この段階ではそれを判定することができない。従ってこの場合には、かかる判断を保留する。

【００７２】そして対話制御部６３は、このような判断処理により、ステップＳＰ１０において、かかる人が新規の人であると判断した場合には、ステップＳＰ１１に

進んで新規学習の開始命令を話者認識部６１及び顔認識部６２に与え、この後ステップＳＰ１２に進んで例えば図１１のように「私はロボットです。よろしくお願いします。」又は「○○さん、今日はいい天気ですね。」などのその人との対話を長引かせる雑談をするための文字列データＤ２を音声合成部６４に送出する。

【００７３】また対話制御部６３は、この後ステップＳＰ１３に進んで話者認識部６１における音響的特徴のデータの収集及び顔認識部６２における顔の形態的特徴のデータの収集が共に十分量に達したか否かを判断し、否定結果を得るとステップＳＰ１２に戻って、この後ステップＳＰ１３において肯定結果を得るまでステップＳＰ１２－ＳＰ１３－ＳＰ１２のループを繰り返す。

【００７４】そして対話制御部６３は、やがて話者認識部６１における音響的特徴のデータの収集及び顔認識部６２における顔の形態的特徴のデータの収集が共に十分量に達することによりステップＳＰ１３において肯定結果を得ると、ステップＳＰ１４に進んで、これら話者認識部６１及び顔認識部６２に新規学習の終了命令を与える。この結果、話者認識部６１において、その音響的特徴のデータが新たなＳＩＤと対応付けられて記憶され、顔認識部６２において、その形態的特徴のデータが新たなＦＩＤと対応付けられて記憶される。

【００７５】また対話制御部６３は、この後ステップＳＰ１５に進んで、話者認識部６１及び顔認識部６２からそれぞれかかるＳＩＤ及びＦＩＤが与えられるのを待ち受け、やがてこれらが与えられると、例えば図１２に示すように、これらをステップＳＰ９において音声認識部６０からの音声認識結果に基づき得られたその人の名前と関連付けてメモリ６５に登録する。そして対話制御部６３は、この後ステップＳＰ２０に進んでその人に対する名前学習処理を終了する。

【００７６】これに対して対話制御部６３は、ステップＳＰ１０において、かかる人が既知の人であると判断した場合には、ステップＳＰ１６に進んで、話者認識部６１及び顔認識部６２がその既知の人を正しく認識できていた場合（すなわち話者認識部６１や顔認識部６２が、関連付け情報としてメモリ６５に格納されたその既知の人に対応するＳＩＤ又はＦＩＤと同じＳＩＤ又はＳＩＤを認識結果として出力していた場合）には、その話者認識部６１又は顔認識部６２に対して追加学習の開始命令を与え、話者認識部６１及び顔認識部６２がその既知の人を正しく認識できなかった場合（すなわち話者認識部６１や顔認識部６２が、関連付け情報としてメモリ６５に格納されたその既知の人に対応するＳＩＤ又はＦＩＤと同じＳＩＤ又はＳＩＤを認識結果として出力していた場合）には、その話者認識部６１又は顔認識部６２に対して訂正学習の開始命令を与える。

【００７７】具体的には、対話制御部６３は、ステップＳＰ９において得られた話者認識部６１からのＳＩＤ

と、最初に顔認識部６２から与えられたＦＩＤとがメモリ６５において同じ名前と関連付けられており、かつステップＳＰ９において音声認識部６０からの音声認識結果に基づき得られた名前がそのＳＩＤ及びＦＩＤが関連付けられた名前であることによりステップＳＰ１０においてその人が既知の人であると判断したときには、話者認識部６１及び顔認識部６２に対してそれぞれ追加学習の開始命令を与える。

【００７８】また話者認識部６３は、ステップＳＰ９において得られた話者認識部６１からのＳＩＤと、最初に顔認識部６２から与えられたＦＩＤとがメモリ６５において異なる名前と関連付けられており、かつステップＳＰ９において音声認識部６０からの音声認識結果に基づき得られた名前がかかるＳＩＤ又はＦＩＤの一方が関連付けられた名前であることによりステップＳＰ１０においてその人が既知の人であると判断したときには、音声認識部６０からの音声認識結果に基づき得られた名前と関連付けられたＳＩＤ又はＦＩＤを出力した一方の話者認識部６１又は顔認識部６２に対して追加学習の開始命令を与え、音声認識部６０からの音声認識結果に基づき得られた名前と関連付けられていないＦＩＤ又はＳＩＤを出力した他方の顔認識部６２又は話者認識部６１に訂正学習の開始命令を与える。

【００７９】そして対話制御部６３は、この後ステップＳＰ１７に進んで、例えば図１３に示すように、「ああ○○さんですね。思い出しましたよ。今日はいい天気ですね。」、「前回はえーと、いつ会いましたっけ。」などのその人との対話を長引かせるための雑談をさせるための文字列データＤ２を音声合成部６４に順次送出し、この後追加学習又は訂正学習に十分な所定時間が経過すると、ステップＳＰ１８に進んで話者認識部６１及び顔認識部６２に対して追加学習又は訂正学習の終了命令を与えた後、ステップＳＰ２０に進んでその人に対する名前学習処理を終了する。

【００８０】他方、対話制御部６３は、ステップＳＰ１０において、かかる人が既知の人であるとも新規の人であるとも判定できないと判断した場合には、ステップＳＰ１９に進んで、例えば図１４に示すように、「ああそうですか。元気ですか。」などの雑談をさせるための文字列データＤ２を音声合成部６４に順次送出する。

【００８１】そしてこの場合には、対話制御部６３は、新規学習、追加学習又は訂正学習の開始命令及びその終了命令を話者認識部６１及び顔認識部６２に与えず（すなわち新規学習、追加学習及び訂正学習のいずれも話者認識部６１及び顔認識部６２に行わせず）、所定時間が経過すると、ステップＳＰ２０に進んでその人に対する名前学習処理を終了する。

【００８２】このようにして対話制御部６３は、音声認識部６０、話者認識部６１及び顔認識部６２の各認識結果に基づいて、人との対話制御や話者認識部６１及び顔認識部６２の動作制御を行うことにより、新規な人の名前を順次学習することができるようになされている。

【００８３】（４）音声認識部６０及び顔認識部６２の具体的構成

次に、上述のような名前学習機能を具現化するための音声認識部６０及び顔認識部６２の具体的構成について説明する。

【００８４】（４－１）音声認識部６０の具体的構成

図１５は、かかる音声認識部６０の具体的構成を示すものである。

【００８５】この音声認識部６０においては、マイクロホン５１からの音声信号Ｓ１ＢをＡＤ（Analog Digital）変換部７０に入力する。ＡＤ変換部７０は、供給されるアナログ信号である音声信号Ｓ１Ｂをサンプリング、量子化し、ディジタル信号である音声データにＡ／Ｄ変換する。この音声データは、特徴抽出部７１に供給される。

【００８６】特徴抽出部７１は、そこに入力される音声データについて、適当なフレームごとに、例えば、ＭＦＣＣ（Mel Frequency Cepstrum Cofficient）分析を行い、その分析の結果得られるＭＦＣＣを、特徴ベクトル（特徴パラメータ）として、マッチング部７２と未登録語区間処理部７６に出力する。なお、特徴抽出部７１では、その後、例えば線形予測係数、ケプストラム係数、線スペクトル対、所定の周波数ごとのパワー（フイルタバンクの出力）等を、特徴ベクトルとして抽出することが可能である。

【００８７】マッチング部７２は、特徴抽出部７１からの特徴ベクトルを用いて、音響モデル記憶部７３、辞書記憶部７４及び文法記憶部７５を必要に応じて参照しながら、マイクロホン５１に入力された音声（入力音声）を、例えば、連続分布ＨＭＭ（Hidden Markov Model）法に基づいて音声認識する。

【００８８】すなわち音響モデル記憶部７３は、音声認識する音声の言語における個々の音素や、音節、音韻などのサブワードについて音響的な特徴を表す音響モデル（例えば、ＨＭＭの他、ＤＰ（Dynamic Programing）マッチングに用いられる標準パターン等を含む）を記憶している。なお、ここでは連続分布ＨＭＭ法に基づいて音声認識を行うことをしているので、音響モデルとしてはＨＭＭ（Hidden Markov Model）が用いられる。

【００８９】辞書記憶部７４は、認識対象の各単位ごとにクラスタリングされた、その単語の発音に関する情報（音響情報）と、その単語の見出しとが対応付けられた単語辞書を認識している。

【００９０】ここで、図１６は、辞書記憶部７４に記憶された単語辞書を示している。

【００９１】図１６に示すように、単語辞書においては、単語の見出しとその音韻系列とが対応付けられており、音韻系列は、対応する単語ごとにクラスタリングさ

れている。図１６の単語辞書では、１つのエントリ（図１６の1行）が、１つのクラスタに相当する。

【００９２】なお、図１６において、見出しはローマ字と日本語（仮名漢字）で表してあり、音韻系列はローマ字で表してある。ただし、音韻系列における「N」は、撥音「ん」を表す。また、図１６では、１つのエントリに１つの音韻系列を記述してあるが、１つのエントリには複数の音韻系列を記述することも可能である。

【００９３】図４に戻り、文法記憶部２６は、辞書記憶部２５の単語辞書に登録されている各単語がどのように連鎖する（つながる）かを記述した文法規則を記憶している。

【００９４】ここで、図１７は、文法記憶部７５に記憶された文法規則を示している。なお、図１７の文法規則は、ＥＢＮＦ（Extended Backus Naur Form）で記述されている。

【００９５】図１７においては、行頭から最初に現れる「;」までが１つの文法規則を表している。また先頭に「$」が付されたアルファベット（列）は変数を表し、「$」が付されていないアルファベット（列）は単語の見出し（図１６に示したローマ字による見出し）を表す。さらに［］で囲まれた部分は省略可能であることを表し、「|」は、その前後に配置された見出しの単語（あるいは変数）のうちのいずれか一方を選択することを表す。

【００９６】従って、図１７において、例えば、第１行（上から１行目）の文法規則「$col＝［Kono | sono］iro wa;」は、変数$colが、「このいろ（色）は」または「そのいろ（色）は」という単語列であることを表す。

【００９７】なお、図１７に示した文法規則においては、変数$silと$garbageが定義されていないが、変数$silは、無音の音響モデル（無音モデル）を表し、変数$garbageは、基本的には、音韻どうしの間での自由な遷移を許可したガーベジモデルを表す。

【００９８】再び図１５に戻り、マッチング部７２は、辞書記憶部７４の単語辞書を参照することにより、音響モデル記憶部７３に記憶されている音響モデルを接続することで、単語の音響モデル（単語モデル）を構成する。さらにマッチング部７２は、幾つかの単語モデルを文法記憶部７５に記憶された文法規則を参照することにより接続し、そのようにして接続された単語モデルを用いて、特徴ベクトルに基づき、連続分布ＨＭＭ法によって、マイクロホン５１に入力された音声を認識する。すなわちマッチング部７２は、特徴抽出部７１が出力する時系列の特徴ベクトルが観測されるスコア（尤度）が最も高い単語モデルの系列を検出し、その単語モデルの系列に対応する単語列の見出しを、音声の認識結果として出力する。

【００９９】より具体的には、マッチング部７２は、接続された単語モデルに対応する単語により接続し、そのようにして接続された単語モデルを用いて、特徴ベクトルに基づき、連続分布ＨＭＭ法によって、マイクロホン５１に入力された音声を認識する。すなわちマッチング部７２は、特徴抽出部７１が出力する時系列の特徴ベクトルが観測されるスコア（尤度）が最も高い単語モデルの系列を検出し、その単語モデルの系列に対応する単語列の見出しを音声認識結果として出力する。

【０１００】より具体的には、マッチング部７２は、接続された単語モデルに対応する単語列について、各特徴ベクトルの出現確率（出力確率）を累積し、その累積値をスコアとして、そのスコアを最も高くする単語列の見出しを音声認識結果として出力する。

【０１０１】以上のようにして出力されるマイクロホン５１に入力された音声認識結果は、文字列データＤ１として対話制御部６３に出力される。

【０１０２】ここで図１７の実施の形態では、第９行（上から９行目）にガーベジモデルを表す変数$garbageを用いた文法規則（以下、適宜、未登録語用規則という）「$pat1＝$color1 $garbage $color2;」があるが、マッチング部７２は、この見登録語用規則が適用された場合には、変数$garbageに対応する音声区間を未登録語の音声区間として検出する。さらに、マッチング部７２は、未登録語用規則が適用された場合における変数$garbageが表すガーベジモデルにおける音韻の遷移としての音韻系列を未登録語の音韻系列として検出する。そしてマッチング部７２は、未登録語用規則が適用された音声認識結果が得られた場合に検出される未登録語の音声区間と音韻系列を未登録語区間処理部７６に供給する。

【０１０３】なお上述の未登録語用規則「$pat1＝$color1 $garbage $color2;」によれば、変数＃color1で表される単語辞書に登録されている単語（列）の音韻系列と、変数$color2で表される単語辞書に登録されている単語（列）の音韻系列との間にある１つの未登録語が検出されるが、この実施の形態においては、発話に複数の未登録語が含まれている場合や、未登録語が単語辞書に登録されている単語（列）間に挟まれていない場合であっても適用可能である。

【０１０４】未登録語区間処理部７６は、特徴抽出部７１から供給される特徴ベクトルの系列（特徴ベクトル系列）を一時記憶する。さらに、未登録語区間処理部７６は、マッチング部７２から未登録語の音声区間と音韻系列を受信すると、その音声区間における音声の特徴ベクトル系列を、一時記憶している特徴ベクトル系列から検出する。そして未登録語区間処理部７６は、マッチング部７２からの音韻系列（未登録語）にユニークなＩＤ（identification）を付し、未登録語の音韻系列と、その音声区間における特徴ベクトル系列とともに、特徴ベクトルバッファ７７に供給する。

【０１０５】特徴ベクトルバッファ７７は、例えば、図

１８に示すように、未登録語区間処理部７６から供給される未登録語のＩＤ、音韻系列及び特徴ベクトル系列を対応付けて一時記憶する。

【０１０６】ここで図１８においては、未登録語に対して１からのシーケンシャルな数時がＩＤとして付されている。従って、例えばいま、特徴ベクトルバッファ７７において、Ｎ個の未登録語のＩＤ、音韻系列及び特徴ベクトル系列が記憶されている場合において、マッチング部７２が未登録語の音声区間と音韻系列を検出すると、未登録語区間処理部７６では、その未登録語に対してＮ＋１がＩＤとして付され、特徴ベクトルバッファ７７では、図１８に点線で示すように、その未登録語のＩＤ、音韻系列及び特徴ベクトル系列が記憶される。

【０１０７】再び図１５に戻り、クラスタリング部７８は、特徴ベクトルバッファ７７に新たに記憶された未登録語（以下、適宜、新未登録語という）について、特徴ベクトルバッファ７７に既に記憶されている他の未登録語（以下、適宜、既記憶未登録語という）それぞれに対するスコアを計算する。

【０１０８】すなわちクラスタリング部７８は、新未登録語を入力音声とし、かつ既記憶未登録語を単語辞書に登録されている単語とみなして、マッチング部７２における場合と同様にして、新未登録語について、各既記憶未登録語に対するスコアを計算する。具体的には、クラスタリング部７８は、特徴ベクトルバッファ７７を参照することで新未登録語の特徴ベクトル系列を認識するとともに、既記憶未登録語の音韻系列にしたがって音響モデルを接続し、その接続された音響モデルから新未登録語の特徴ベクトル系列が観測される尤度としてのスコアを計算する。

【０１０９】なお、音響モデルは、音響モデル記憶部７３に記憶されているものが用いられる。

【０１１０】クラスタリング部７８は、同様にして、各既記憶未登録語について、新未登録語に対するスコアも計算し、そのスコアによってスコアシート記憶部７９に記憶されたスコアシートを更新する。

【０１１１】さらにクラスタリング部７８は、更新したスコアシートを参照することにより、既に求められている未登録語（既記憶未登録語）をクラスタリングしたクラスタの中から、新未登録語を新たなメンバとして加えるクラスタを検出する。さらにクラスタリング部７８は、新未登録語を検出したクラスタの新たなメンバとし、そのクラスタをそのクラスタのメンバに基づいて分割し、その分割結果に基づいて、スコアシート記憶部７９に記憶されているスコアシートを更新する。

【０１１２】スコアシート記憶部７９は、新未登録語についての既記憶未登録語に対するスコアや、既記憶未登録語についての新未登録語に対するスコア等が登録されたスコアシートを記憶する。

【０１１３】ここで、図１９は、スコアシートを示している。

【０１１４】スコアシートは、未登録語の「ＩＤ」、「音韻系列」、「クラスタナンバ」、「代表メンバＩＤ」及び「スコア」が記述されたエントリで構成される。

【０１１５】未登録語の「ＩＤ」と「音韻系列」としては、特徴ベクトルバッファ７７に記憶されたものと同一のものがクラスタリング部７８によって登録される。「クラスタナンバ」は、そのエントリの未登録語がメンバとなっているクラスタを特定するための数字で、クラスタリング部７８によって付され、スコアシートに登録される。「代表ナンバＩＤ」は、そのエントリの未登録語がメンバとなっているクラスタを代表する代表メンバとしての未登録のＩＤであり、この代表メンバＩＤによって、未登録語がメンバとなっているクラスタの代表メンバを認識することができる。なお、クラスタの代表メンバは、クラスタリング部２９によって求められ、その代表メンバのＩＤがスコアシートの代表メンバＩＤに登録される。「スコア」は、そのエントリの未登録語についての他の未登録語それぞれに対するスコアであり、上述したように、クラスタリング部７８によって計算される。

【０１１６】例えば、いま、特徴ベクトルバッファ７７において、Ｎ個の未登録語のＩＤ、音韻系列及び特徴ベクトル系列が記憶されているとすると、スコアシートには、そのＮ個の未登録語のＩＤ、音韻系列、クラスタナンバ、代表ナンバＩＤ及びスコアが登録されている。

【０１１７】そして特徴ベクトルバッファ７７に、新未登録語のＩＤ、音韻系列、および特徴ベクトル系列が新たに記憶されると、クラスタリング部７８では、スコアシートが図１９において点線で示すように更新される。

【０１１８】すなわちスコアシートには、新未登録語のＩＤ、音韻系列、クラスタナンバ、代表メンバＩＤ、新未登録語についての既記憶未登録語それぞれに対するスコア（図１９におけるスコアｓ（Ｎ+１，１）、ｓ（２、Ｎ+１）、…ｓ（Ｎ+１、Ｎ）が追加される。さらにスコアシートには、既記憶未登録語それぞれについての新未登録語に対するスコア（図１９におけるｓ（Ｎ+１，１）、ｓ（２、Ｎ+１）、…ｓ（Ｎ+１、Ｎ））が追加される。さらに後述するように、スコアシートにおける未登録語のクラスタナンバと代表メンバＩＤが必要に応じて変更される。

【０１１９】なお、図１９の実施の形態においては、ＩＤがｉの未登録語（の発話）についての、ＩＤがｊの未登録語（の音韻系列）に対するスコアを、ｓ（ｉ、ｊ）として表してある。

【０１２０】またスコアシート（図１９）には、ＩＤがｉの未登録語（の発話）についての、ＩＤがｉの未登録語（の音韻系列）に対するスコアｓ（ｉ、ｊ）も登録される。ただし、このスコアｓ（ｉ、ｊ）は、マッチング部７２において、未登録語の音韻系列が検出されるときに計算されるため、クラスタリング部７８で計算する必

要はない。

【０１２１】再び図１５に戻り、メンテナンス部８０は、スコアシートに記憶部７９における更新後のスコアシートに基づいて、辞書記憶部７４に記憶された単語辞書を更新する。

【０１２２】ここで、クラスタの代表メンバは、次のように決定される。すなわち、例えば、クラスタのメンバとなっている未登録語のうち、他の未登録語それぞれに＊ついてのスコアの総和（その他、例えば、総和を他の未登録語の数で除算した平均値でも良い）を最大にするものがそのクラスタの代表メンバとされる。従って、この場合、クラスタに属するメンバのメンバＩＤをｋで表すこととすると、次式

【０１２３】

【数１】

$$K=\max_{k}\left\{\Sigma s(k^{3},k)\right\} \quad \cdots\cdots (1)$$

【０１２４】で示される値ｋ（∈ｋ）をＩＤとするメンバが代表メンバとされることになる。

【０１２５】ただし、（１）式において、$\max_{k}\{\}$は、{}内の値を最大にするｋを意味する。またk^{3}は、ｋと同様に、クラスタに属するメンバのＩＤを意味する。さらに、Σは、k^{3}をクラスタに属するメンバすべてのＩＤに亘って変化させての総和を意味する。

【０１２６】なお上述のように代表メンバを決定する場合、クラスタのメンバが１または２つの未登録語であるときには、代表メンバを決めるにあたってスコアを計算する必要はない。すなわちクラスタのメンバが１つの未登録語である場合には、その１つの未登録語が代表メンバとなり、クラスタのメンバが２つの未登録語である場合には、その２つの未登録語のうちのいずれを代表メンバとしても良い。

【０１２７】また代表メンバの決定方法は、上述したものに限定されるものではなく、その他、例えばクラスタのメンバとなっている未登録語のうち、他の未登録語それぞれとの特徴ベクトル空間における距離の総和を最小にするもの等をそのクラスタの代表メンバとすることも可能である。

【０１２８】以上のように構成される音声認識部６０では、マイクロホン５１に入力された音声を認識する音声認識処理と、未登録語に関する未登録語処理が図２０に示す音声認識処理手順ＲＴ２に従って行われる。

【０１２９】実際上、音声認識部６０では、人が発話を行うことにより得られた音声信号Ｓ１Ｂがマイクロホン５１からＡＤ変換部７０を介して音声データとされて特徴抽出部７１に与えられるとこの音声認識処理手順ＲＴ２がステップＳＰ３０において開始される。

【０１３０】そして続くステップＳＰ３１において、特徴抽出部７１が、その音声データを所定のフレーム単位で音響分析することにより特徴ベクトルを抽出し、その特徴ベクトルの系列をマッチング部７２及び未登録語区間処理部７６に供給する。

【０１３１】マッチング部７６は、続くステップＳ３２において、特徴抽出部７１からの特注オベクトル系列について、上述したようにスコア計算を行い、この後ステップＳ３３において、スコア計算の結果得られるスコアに基づいて、音声認識結果となる単語列の見出しを求めて出力する。

【０１３２】さらにマッチング部７２は、続くステップＳ３４において、ユーザの音声に未登録語が含まれていたかどうかを判定する。

【０１３３】ここで、このステップＳ３４において、ユーザの音声に未登録語が含まれていないと判定された場合、すなわち上述の未登録語用規則「$pat1＝$color1 $garbage $color2；」が適用されずに音声認識結果が得られた場合、ステップＳ３５に進んで処理が終了する。

【０１３４】これに対してステップＳ３４において、ユーザの音声に未登録語が含まれていると判定された場合、すなわち未登録語用規則「$pat1＝$color1 $garbage $color2；」が適用されて音声認識結果が得られた場合、マッチング部２３は、続くステップＳ３５において、未登録語用規則の変数$garbageに対応する音声区間を未登録語の音声区間として検出するとともに、その変数$garbageが表すガーベジモデルにおける音韻の遷移としての音韻系列を未登録語の音韻系列として検出し、その未登録語の音声区間と音韻系列を未登録語区間処理部７６に供給して、処理を終了する（ステップＳＰ３６）。

【０１３５】一方、未登録語機関処理部７６は、特徴抽出部７１から供給される特徴ベクトル系列を一時記憶しており、マッチング部７２から未登録語の音声区間と音韻系列が供給されると、その音声区間における音声の特徴ベクトル系列を検出する。さらに未登録語区間処理部７６は、マッチング部７２からの未登録語（の音韻系列）にＩＤを付し、未登録語の音韻系列と、その音声区間における特徴ベクトル系列とともに、特徴ベクトルバッファ７７に供給する。

【０１３６】以上のようにして、特徴ベクトルバッファ７７に新たな未登録語（新未登録語）のＩＤ、音韻系列及び特徴ベクトル系列が記憶されると、この後、未登録語の処理が図２１に示す未登録語処理手順ＲＴ３に従って行われる。

【０１３７】すなわち音声認識部６０においては、上述のように特徴ベクトルバッファ７７に新たな未登録語（新未登録語）のＩＤ、音韻系列及び特徴ベクトル系列が記憶されるとこの未登録語処理手順ＲＴ３がステップ

ＳＰ４０において開始され、まず最初にステップＳ４１において、クラスタリング部７８が、特徴ベクトルバッファ７７から新未登録語のＩＤと音韻系列を読み出す。

【０１３８】次いでステップＳ４２において、クラスタリング部７８が、スコアシート記憶部３０のスコアシートを参照することにより、既に求められている（生成されている）クラスタが存在するかどうかを判定する。

【０１３９】そしてこのステップＳ４２において、すでに求められているクラスタご存在しないと判定された場合、すなわち新未登録語が初めての未登録語であり、スコアシートに既記憶未登録語のエントリが存在しない場合には、ステップＳ４３に進み、クラスタリング部７８が、その新未登録語を代表メンバとするクラスタを新たに生成し、その新たなクラスタに関する情報と、親身登録語に関する情報とをスコアシート記憶部７９のスコアシートに登録することにより、スコアシートを更新する。

【０１４０】すなわちクラスタリング部７８は、特徴ベクトルバッファ７７から読み出した新未登録語のＩＤおよび音韻系列をスコアシート（図１９）に登録する。さらにクラスタリング部７８は、ユニークなクラスタナンバを生成し、新未登録語のクラスタナンバとしてスコアシートに登録する。またクラスタリング部７８は、新未登録語のＩＤをその新未登録語の代表ナンバＩＤとして、スコアシートに登録する。従ってこの場合は、新未登録語は、新たなクラスタの代表メンバとなる。

【０１４１】なお、いまの場合、新未登録語とのスコアを計算する既記憶未登録語が存在しないため、スコアの計算は行われない。

【０１４２】かかるステップＳ４３の処理後は、ステップＳ５２に進み、メンテナンス部８０は、ステップＳ４３で更新されたスコアシートに基づいて、辞書記憶部７４の単語辞書を更新し、処理を終了する（ステップＳＰ５４）。

【０１４３】すなわち、いまの場合、新たなクラスタが生成されているので、メンテナンス部３１は、スコアシートにおけるクラスタナンバを参照し、その新たに生成されたクラスタを認識する。そしてメンテナンス部８０は、そのクラスタに対応するエントリを辞書記憶部７４の単語辞書に追加し、そのエントリの音韻系列として、新たなクラスタの代表メンバの音韻系列、つまりいまの場合は、新未登録語の音韻系列を登録する。

【０１４４】一方、ステップＳ４２において、すでに求められているクラスタが存在すると判定された場合、すなわち新未登録語が初めての未登録語ではなく、従ってスコアシート（図１９）に、既記憶未登録語のエントリ（行）が存在する場合、ステップＳ４４に進み、クラスタリング部７８は、新未登録語について、各既記憶未登録語それぞれに対するスコアを計算すると共に、各既記憶未登録語それぞれについて、新未登録語に対するスコアを計算する。

【０１４５】すなわち、例えば、いま、ＩＤが１乃至Ｎ個の既記憶未登録語が存在し、新未登録語のＩＤをＮ＋１とすると、クラスタリング部７８では、図１９において点線で示した部分の新未登録語についてのＮ個の既記憶未登録語それぞれに対するスコアｓ（Ｎ＋１、１）、ｓ（Ｎ＋１、２）…、ｓ（Ｎ、Ｎ＋１）と、Ｎ個の既記憶未登録語それぞれについての新未登録語に対するスコアｓ（１、Ｎ＋１）、ｓ（２、Ｎ＋１）…、ｓ（Ｎ、Ｎ＋１）が計算される。なおクラスタリング部７８において、これらのスコアを計算するにあたっては、新未登録語とＮ個の既記憶未登録語それぞれの特徴ベクトル系列が必要となるが、これらの特徴ベクトル系列は、特徴ベクトルバッファ２８を参照することで認識される。

【０１４６】そしてクラスタリング部７８は、計算したスコアを新未登録語のＩＤ及び音韻系列とともにスコアシート（図１９）に追加し、ステップＳ４５に進む。

【０１４７】ステップＳ４５では、クラスタリング部７８はスコアシート（図１９）を参照することにより、新未登録語についてのスコアｓ（Ｎ＋１、ｉ）（ｉ＝１、２、…、Ｎ）を最も高く（大きく）する代表メンバを有するクラスタを検出する。即ち、クラスタリング部７８は、スコアシートの代表メンバＩＤを参照することにより、代表メンバとなっている既記憶未登録語を認識し、さらにスコアシートのスコアを参照することで、新未登録語についてのスコアを最も高くする代表メンバとしての既記憶未登録語を検出する。そしてクラスタリング部７８は、その検出した代表メンバとしての既記憶未登録語のクラスタナンバのクラスタを検出する。

【０１４８】その後、ステップＳ４６に進み、クラスタリング部２９は、新未登録語をステップＳ４５で検出したクラスタ（以下、適宜、検出クラスタという）のメンバに加える。すなわちクラスタリング部７８は、スコアシートにおける新未登録語のクラスタナンバとして、検出クラスタの代表メンバのクラスタナンバを書き込む。

【０１４９】そしてクラスタリング部７８は、ステップＳ４７において、検出クラスタを例えば２つのクラスタに分割するクラスタ分割処理を行い、ステップＳ４８に進む。ステップＳ４８では、クラスタリング部７８は、ステップＳ４７のクラスタ分割処理によって、検出クラスタを２つのクラスタに分割することができたかどうか判定し、分割することができた判定した場合、ステップＳ４９に進む。ステップＳ４９では、クラスタリング部７８は、検出クラスタの分割により得られる２つのクラスタ（この２つのクラスタを、以下、適宜、第１の子クラスタと第２の子クラスタという）同士の間のクラスタ間距離を求める。

【０１５０】ここで、第１及び第２の子クラスタ同士間のクラスタ間距離とは、例えば次のように定義される。

【０１５１】すなわち第１の子クラスタと第２の子クラ

スタの両方の任意のメンバ（未登録語）のＩＤを、ｋで表すとともに、第１と第２の子クラスタの代表メンバ（未登録語）のＩＤを、それぞれｋ１またはｋ２で表す＊

＊こととすると、次式

【０１５２】

【数２】

$$D(k1,k2)=\max val_{k}\left\{abs(\log(s(k,k1))-\log(s(k,k2)))\right\} \quad \cdots\cdots (2)$$

【０１５３】で表される値Ｄ（ｋ１，ｋ２）を第１と第２の子クラスタ同士の間のクラスタ間距離とする。

【０１５４】ただし、（２）式において、abs（）は、（）内の値の絶対値を表す。また、maxval$_k$ {}は、ｋを変えて求められる{}内の値の最大値を表す。またlogは、自然対数又は常用対数を表す。

【０１５５】いま、ＩＤがｉのメンバをメンバ＃Ｉと表すこととすると、（２）式におけるスコアの逆数１／ｓ（ｋ，ｋ１）は、メンバ＃ｋと代表メンバｋ１との距離に相当し、スコアの逆数１／ｓ（ｋ，ｋ２）は、ﾒﾝﾊﾞ＃ｋと代表メンバｋ２との距離に相当する。従って、（２）式によれば、第１と第２の子クラスタのメンバのうち、第１の子クラスタの代表メンバ＃ｋ１との距離と、第２の子クラスタの代表メンバ＃ｋ２との差の最大値が、第１と第２の子クラスタ同士の間の子クラスタ間距離とされることになる。

【０１５６】なおクラスタ間距離は、上述したものに限定されるものではなく、その他、例えば、第１の子クラスタの代表メンバと、第２の子クラスタの代表メンバとのＤＰマッチングを行うことにより、特徴ベクトル空間における距離の積算値を求め、その距離の積算値を、クラスタ間距離とすることも可能である。

【０１５７】ステップＳ４９の処理後は、ステップＳ５０に進み、クラスタリング部７８は、第１と第２の子クラスタ同士のクラスタ逢間距離が、所定の閾値ξより大である（あるいは、閾値ξ以上である）かどうかを判定する。

【０１５８】ステップＳ５０において、クラスタ間距離が所定の閾値ξより大であると判定された場合、すなわち検出クラスタのメンバとしての複数の未登録後が、その音響的特徴からいって、２つのクラスタにクラスタリングすべきものであると考えられる場合、ステップＳ５１に進み、クラスタリング部７８は、第１と第２の子クラスタをスコアシート記憶部７９のスコアシートに登録する。

【０１５９】すなわちクラスタリング部７８は、第１と第２の子クラスタにユニークなクラスタナンバを割り当て、検出クラスタのメンバのうち、第１の子クラスタにクラスタリングされたもののクラスタナンバを第１の子クラスタのクラスタナンバにすると共に、第２の子クラスタにクラスタリングされたもののクラスタナンバを第２の子クラスタのクラスタナンバにするように、スコアシートを更新する。

【０１６０】さらにクラスタリング部７８は、第１の子クラスタにクラスタリングされたメンバの代表メンバＩＤを第１の子クラスタの代表メンバのＩＤにすると共に、第２の子クラスタにクラスタリングされたメンバの代表メンバＩＤを第２の子クラスタの代表メンバのＩＤにするように、スコアシートを更新する。

【０１６１】なお、第１と第２の子クラスタのうちいずれか一方には、検出クラスタのクｒスタナンバを割り当てるようにすることが可能である。

【０１６２】クラスタリング部７８が以上のようにして第１と第２の子クラスタをスコアシートに登録すると、ステップＳ５１からＳ５２に進み、メンテナンス部８０が、スコアシートに基づいて、辞書記憶部７４の単語辞書を更新し、処理を終了する（ステップＳＰ５４）。

【０１６３】すなわち、いまの場合、検出クラスタが第１と第２の子クラスタに分割されたため、メンテナンス部８０は、まず単語辞書における検出クラスタに対応するエントリを削除する。さらにメンテナンス部８０は、第１と第２の子クラスタそれぞれに対応する２つのエントリを単語辞書に追加し、第１の子クラスタに対応するエントリの音韻系列として、その第１の子クラスタの代表メンバの音韻系列を登録すると共に、第２の子クラスタに対応するエントリの音韻系列として、その第２の子クラスタの代表メンバの音韻系列を登録する。

【０１６４】一方、ステップＳ４８において、ステップＳ４７のクラスタ分割処理によって、検出クラスタを２つのクラスタに分割することができなかったと判定された場合、又はステップＳ５０において、第１と第２の子クラスタのクラスタ間距離が所定の閾値ξより大でないと判定された場合、従って、検出クラスタのメンバとしての複数の未登録後の音響的特徴が第１と第２の子クラスタにクラスタリングするほど似ていないものではない場合）、ステップＳ５３に進み、クラスタリング部７８は、検出クラスタの新たな代表メンバを求め、スコアシートを更新する。

【０１６５】すなわちクラスタリング部７８は、新未登録後をメンバとして加えた検出クラスタの各メンバについて、スコアシート記憶部７９のスコアシートを参照することにより、（１）式の計算に必要なスコアｓ（k^3，ｋ）を認識する。さらに、クラスタリング７８は、その認識したスコアｓ（k^3，ｋ）を用い、（１）式に基づき、検出クラスタの新たな代表メンバとなるメンバのＩＤを求める。そしてクラスタリング部７８は、スコアシート（図１９）における検出クラスタの各メンバの代表メンバＩＤを、検出クラスタの新たな代表メンバのＩＤに書き換える。

【０１６６】その後、ステップＳ５２に進み、メンテナ

ンス部８０が、スコアシートに基づいて辞書記憶部７４の単語辞書を更新し、処理を終了する（ステップＳＰ５４）。

【０１６７】すなわち、いまの場合、メンテナンス部８０は、スコアシートを参照することにより、検出クラスタの新たな代表メンバを認識し、さらにそのダ表メンバの音韻系列を認識する。そしてメンテナンス部８０は、単語辞書における検出クラスタに対応するエントリの音韻系列を、検出クラスタの新たな代表メンバの音韻系列に変更する。

【０１６８】ここで、図２１のステップＳＰ４７のクラスタ分割処理は、図２２に示すクラスタ分割処理手順ＲＴ４に従って行われる。

【０１６９】すなわち音声認識処理部６０では、図２２のステップＳＰ４６からステップＳＰ４７に進むとこのクラスタ分割処理手順ＲＴ４をステップＳＰ６０において開始し、まず最初にステップＳ６１において、クラスタリング部７８が、新未登録後がメンバとして加えられた検出クラスタから、まだ選択していない任意の２つのメンバの組み合わせを選択し、それぞれを仮の代表メンバとする。ここで、この２つの仮の代表メンバを、以下、適宜、第１の仮代表メンバと第２の仮代表メンバという。

【０１７０】そして、続くステップＳ６２において、クラスタリング部７８は、第１の仮代表メンバ及び第２の仮代表メンバをそれぞれ代表メンバとすることができるように、検出クラスタのメンバを２つのクラスタに分割することができるかどうかを判定する。

【０１７１】ここで、第１又は第２の仮代表メンバを代表メンバとすることができるかどうかは（１）式の計算を行う必要があるが、この計算に用いられるスコアｓ（ｋ'，ｋ）は、スコアシートを参照することで認識される。

【０１７２】ステップＳ６２において、第１の仮代表メンバ及び第２の仮代表メンバをそれぞれ代表メンバとすることができるように、検出クラスタのメンバを２つのクラスタに分割することができないと判定された場合、ステップＳ６２をスキップして、ステップＳ６４に進む。

【０１７３】また、ステップＳ６２において、第１の仮代表メンバと、第２の仮代表メンバをそれぞれ代表メンバとすることができるように、検出クラスタのメンバを２つのクラスタに分割することができると判定された場合、ステップＳ６３に進み、クラスタリング部７８は、第１の仮代表メンバと、第２の仮代表メンバがそれぞれ代表メンバとなるように、検出クラスタのメンバを２つのクラスタに分割し、その分割後の２つのクラスタの組を、検出クラスタの分割結果となる第１及び第２の子クラスタの候補（以下、適宜、候補クラスタの組という）として、ステップＳ６４に進む。

【０１７４】ステップＳ６４では、クラスタリング部７８は、検出クラスタのメンバの中で、まだ第１と第２の仮代表メンバの組として選択していない２つのメンバの組があるかどうかを判定し、あると判定した場合、ステップＳ６１に戻り、まだ第１と第２の仮代表メンバの組として選択していない検出クラスタの２つのメンバの組が選択され、以下、同様の処理が繰り返される。

【０１７５】またステップＳ６４において、第１と第２の仮代表メンバの組として選択していない検出クラスタの２つのメンバの組がないと判定された場合、ステップＳ６５に進み、クラスタリング部７８は、候補クラスタの組が存在するかどうかを判定する。

【０１７６】ステップＳ６５において、候補クラスタの組が存在しないと判定された場合、ステップＳ６６をスキップして、リターンする。この場合は、図２１のステップＳ４８において、検出クラスタを分割することができなかったと判定される。

【０１７７】一方、ステップＳ６５において、候補クラスタの組が存在すると判定された場合、ステップＳ６６に進み、クラスタリング部７８は、候補クラスタの組が複数存在するときには、各候補クラスタの組の２つのクラスタ同士の間のクラスタ間距離を求める。そして、クラスタリング部７８は、クラスタ間距離が最小の候補クラスタの組を求め、その候補クラスタの組を検出クラスタの分割結果をして、すなわち第１と第２の子クラスタとして、リターンする。なお、候補クラスタの組が１つだけの場合は、その候補クラスタの組がそのまま第１と第２の子クラスタとされる。

【０１７８】この場合は、図２１のステップＳ４８において、検出クラスタを分割することができたと判定される。

【０１７９】以上のように、クラスタリング部７８において、既に求められている未登録語をクラスタリングしたクラスタの中から、新未登録語を新たなメンバとして加えるクラスタ（検出クラスタ）を検出し、新未登録語をその検出クラスタの新たなメンバとして、検出クラスタをその検出クラスタのメンバに基づいて分割するようにしたので、未登録語をその音響的特徴が近似しているもの同士に容易にクラスタリングすることができる。

【０１８０】さらにメンテナンス部８０において、そのようなクラスタリング結果に基づいて単語辞書を更新するようにしたので、単語辞書の大規模化を避けながら、未登録語の単語辞書への登録を容易に行うことができる。

【０１８１】また、例えば、仮に、マッチング部７２において、未登録語の音声区間の検出を誤ったとしても、そのような未登録語は、検出クラスタの分割によって、音声区間が正しく検出された未登録語とは別のクラスタにクラスタリングされる。そして、このようなクラスタに対応するエントリが単語辞書に登録されることになる

が、このエントリの音韻系列は正しく検出されなかった音声区間に対応するものとなるから、その後の音声認識において大きなスコアを与えることはない。従って、仮に、未登録語の音声区間の検出を誤ったとしても、その誤りはその後の音声認識にはほとんど影響しない。

【０１８２】ここで、図２３は、未登録語の発話を行って得られたクラスタリング結果を示している。なお、図２３においては、各エントリ（各行）が１つのクラスタを表している。また、図２３の左欄は、各クラスタの代表メンバ（未登録語）の音韻系列を表しており、図２３の右欄は、各クラスタのメンバとなっている未登録語の発話内容と数を表している。

【０１８３】すなわち図２３において、例えば第１行のエントリは、未登録語「風呂」の１つの発話だけがメンバとなっているクラスタを表しており、その代表メンバの音韻系列は、「doroa：」（ドロアー）になっている。また、例えば第２行のエントリは、未登録語「風呂」の３つの発話がメンバとなっているクラスタを表しており、その代表メンバの音韻系列は、「kuro」（クロ）になっている。

【０１８４】さらに、例えば第７行のエントリは、未登録語「本」の４つの発話がメンバとなっているクラスタを表しており、その代表メンバの音韻系列は、「NhoNde：su」（ンホンテース）になっている。また、例えば第８行のエントリは、未登録語「オレンジ」の１つの発話と、未登録語「本」の１９の発話がメンバとなっているクラスタを表しており、その代表メンバの音韻系列は、「ohoN」（オホン）になっている。他のエントリも同様のことを表している。

【０１８５】図２３によれば、同一の未登録語の発話について、良好にクラスタリングされていることが分かる。

【０１８６】なお、図２３の第８行のエントリにおいては、未登録語「オレンジ」の１つの発話と、未登録語「本」の１９の発話が、同一のクラスタにクラスタリングされている。このクラスタはそのメンバとなっている発話から、未登録語「本」のクラスタとなるべきであると考えられるが、未登録語「オレンジ」の発話も、そのクラスタのメンバとなっている。しかしながらこのクラスタも、その後に未登録語「本」の発話がさらに入力されていくと、クラスタ分割され、未登録語「本」の発話だけをメンバとするクラスタと、未登録語「オレンジ」の発話だけをメンバとするクラスタにクラスタリングされると考えられる。

【０１８７】（４－２）顔認識部６２の具体的構成

次に、顔認識部６２の具体的構成について説明する。

【０１８８】図２４及び図２５に示すように、顔認識部６２は、動的に変化する環境下で一定時間内に応答する＊ことができるが、ＣＣＤカメラ５０（図５）から与えられる画像信号Ｓ１Ａに基づく画像内から顔パターンを抽出する顔抽出処理部９０と、抽出された顔パターンを基に顔を認識する顔認識処理部９１から構成される。本実施の形態では、顔パターンを抽出する顔抽出処理に「ガボア・フィルタリング（Gabor Filtering）」を採用し、また、顔パターンから顔を認識する顔認識処理には「サポート・ベクタ・マシーン（Support Vector Machine：ＳＶＭ）」を採用している。

【０１８９】この顔認識部６２は、顔パターンを顔認識処理部９１が学習する学習段階と、学習されたデータを基に、画像信号Ｓ１Ａから抽出された顔パターンを認識する認識段階を持つ。

【０１９０】図２４には、顔認識部６２の学習段階の構造を、また図２５には、顔認識部６２の認識段階の構成をそれぞれ示している。

【０１９１】学習段階においては、図２４に示すように、ＣＣＤカメラ５０（図５）から入力されたユーザの撮像画像をガボア・フィルタでなる顔抽出処理部９０で顔抽出した結果がサポート・ベクタ・マシーンでなる顔認識処理部９１に投入される。顔認識処理部９１では、外部から供給される学習用のデータすなわち教師データを用いて、暫定的な識別関数を得る。

【０１９２】また、識別段階においては、図２５に示すように、ＣＣＤカメラ５０から供給される画像信号Ｓ１Ａに基づく画像内の人の顔を顔抽出処理部９０で顔抽出した結果が顔認識処理部９１に投入される。顔認識処理部９１では、暫定的に得られた識別関数をさまざまなデータベース上の画像に試して顔の検出を行う。そして、検出に成功したものを顔データとして出力する。また検出に失敗したものを非顔データとして学習データに追加して、さらに学習をし直す。

【０１９３】以下、顔抽出処理部９０におけるガボア・フィルタリング処理と、顔認識処理部９１におけるサポート・ベクタ・マシーンについて、それぞれ詳細に説明する。

【０１９４】（４－２－１）ガボア・フィルタリング処理

人間の視覚細胞には、ある特定の方位に対して選択性を持つ細胞が存在することが既に判っている。これは、垂直の線に対して反応する細胞と、水平の線に反応する細胞で構成される。ガボア・フィルタリングは、これと同様に、方位選択性を持つ複数のフィルタで構成される空間フィルタである。

【０１９５】ガボア・フィルタは、ガボア関数によって空間表現される。ガボア関数ｇ（ｘ、ｙ）は、次式

【０１９６】

【数３】

$$g(x,y)=s(x,y)w_r(x,y) \quad \cdots\cdots (3)$$

33

【０１９７】に示すように、コサイン成分からなるキャリアｓ（ｘ、ｙ）と、２次元ガウス分析状のエンベロープｗ，（ｘ、ｙ）とで構成される。

【０１９８】キャリアｓ（ｘ、ｙ）は、複数関数を用いて、下式（４）のように表現される。ここで、座標値＊

$$s(x,y)=\exp(j(2\pi(u_0x+v_0y)+P))$$

【０２０１】に示すキャリアは、次式

【０２０２】※

$$Re(s(x,y))=\cos(2\pi(u_0x+v_0y)+P)$$

$$Im(s(x,y))=\sin(2\pi(u_0x+v_0y)+P) \quad \cdots\cdots (5)$$

【０２０３】に示すように、実数成分Ｒｅ（ｓ（ｘ、ｙ）と虚数成分Ｉｍ（ｓ（ｘ、ｙ））に分離することができる。

【０２０４】一方、２次元ガウス分布からなるエンベロ★

$$w_r(x,y)=K\exp(-\pi(a^2(x-x_0)_r^2+b^2(y-y_0)_r^2)) \quad \cdots\cdots (6)$$

【０２０６】のように表現される。

【０２０７】ここで、座標軸（x_0、y_0）はこの関数のピークであり、定数ａ及びｂはガウス分布のスケール・☆

$$(x-x_0)_r=(x-x_0)\cos\theta+(y-y_0)\sin\theta$$

$$(y-y_0)_r=-(x-x_0)\sin\theta+(y-y_0)\cos\theta \quad \cdots\cdots (7)$$

【０２０９】に示すような回転操作を意味する。

【０２１０】従って、上述の（４）式及び（６）式より、ガボア・フィルタは、次式

◆

$$g(x,y)=K\exp(-\pi(a^2(x-x_0)_r^2+b^2(y-y_0)_r^2))\exp(j(2\pi(u_0x+u_0y)+P)) \quad \cdots\cdots (8)$$

【０２１２】に示すような空間関数として表現される。

【０２１３】本実施の形態に係る顔抽出処理部９０は、８種類の方向と３通りの周波数は採用して、合計２４個のガボア・フィルタを用いて顔抽出処理を行う。

【０２１４】ガボア・フィルタのレスポンスは、G_iを＊

$$J_i(x,y)=G_i(x,y)\oplus I(x,y) \quad \cdots\cdots (9)$$

【０２１６】で表される。この（９）式の演算は、実際には高速フーリエ変換を用いて高速化することができる。

【０２１７】作成したガボア・フィルタの性能を調べる※

$$H(x,y)=\sum_{i=1}^{0}a_iJ_i(x,y) \quad \cdots\cdots (10)$$

【０２１９】のように表される。

【０２２０】そして、入力画像Ｉと再構築された画像Ｈとのエラー Ｅは、次式

★

$$E=\frac{1}{2}||I(x,y)-H(x,y)||^2=\frac{1}{2}\sum_{x,y}(I(x,y)-H(x,y))^2 \quad \cdots\cdots (11)$$

【０２２２】のように表される。

34

＊（u_0、v_0）は空間周波数を表し、またＰはコサイン成分の位相を表す。

【０１９９】ここで、次式

【０２００】

【数４】

※【数５】

★ーブは、次式

【０２０５】

【数６】

☆パラメータである。また、添え字ｒは、次式

【０２０８】

【数７】

◆【０２１１】

【数８】

＊ｉ番目のガボア・フィルタとし、ｉ番目のガボアの結果（Gabor Jet）をJ_iとし、入力イメージをＩとし、すると、次式

【０２１５】

【数９】

※ためには、フィルタリングして得られた画素を再構築することによって行う。再構築されたイメージＨは、次式

【０２１８】

【数１０】

★【０２２１】

【数１１】

【０２２３】このエラーＥを最小にするような最適なａ

35

を求めることにより再構築することができる。

【０２２４】（４－２－２）サポート・ベクタ・マシーン本実施の形態では、顔認識処理部９１における顔認識に関して、パターン認識の分野で最も学習汎化能力が高いとされるサポート・ベクタ・マシーン（ＳＶＭ）を用いて該当する顔か否かの識別を行う。

【０２２５】ＳＶＭ自体に関しては、例えばB. sholkopf外著の報告（B. Sholkopf、C. Burges、A. Smola、"Advance in Kernel Support Vector Learning"、The MIT Press、1999.）を挙げることができる。本願出願人が行った予備実験の結果からは、ＳＶＭによる顔認識方法は、主成分分析（ＰＣＡ）やニューラル・ネットワークを用＊

$$f(x)=\sum_{i}^{n} w_i x_i + b$$

【０２２９】で与えられる識別関数ｆ（ｘ）を求めることである。

【０２３０】ここで、ＳＶＭの学習用の教師ラベルを次※

$$y=(y1, y2, \cdots\cdots, yn)$$

【０２３２】のようにおく。

【０２３３】すると、ＳＶＭにおける顔パターンの認識を次式 ★

$$y_i(w' x_i + b) \geq 1$$

【０２３５】に示す制約条件の下での重み因子ｗの二乗の最小化する問題としてとらえることができる。

【０２３６】このような制約のついた問題は、ラグランジュの未定定数法を用いて解くことができる。すなわ ☆

$$L(w, b, a)=\frac{1}{2}||w||^2-\sum_{i}^{l} a_i (y_i ((x_i' w + b)-1)) \quad \cdots\cdots (15)$$

【０２３８】に示すラグランジュをまず導入し、次いで、次式 ◆

$$\frac{\partial L}{\partial b}=\frac{\partial L}{\partial w}=0 \quad \cdots\cdots (16)$$

【０２４０】に示すように、ｂ、ｗの各々について偏微分する。

【０２４１】この結果、ＳＶＭにおける顔パターンの識＊

$$\max \sum ai-\frac{1}{}\sum aiaiyiyi' xj$$

制約条件：$ai \geqq 0, \sum aiyi=0$ …… （１７）

【０２４３】に示す２次計画問題としてとらえることができる。

【０２４４】特徴空間の次元数が、訓練サンプルの数よりも少ない場合は、スクラッチ変数$\xi \geqq 0$を導入して、※

$$y_i (w' x_i + b) \geqq 1 - \xi_i$$

36

＊いる手法に比べ、良好な結果を示すことが判っている。

【０２２６】ＳＶＭは、識別関数に線形識別器（パーセプトロン）を用いた学習機械であり、カーネル関数を使うことで非線形空間に拡張することができる。また識別関数の学習では、クラス間分離のマージンを最大にとるように行われ、その解は２次数理計画法を解くことで得られるため、グローバル解に到達できることを理論的に保証することができる。

【０２２７】通常、パターン認識の問題は、テスト・サンプルｘ＝（ｘ１、ｘ２…。ｘｎ）に対して、次式

【０２２８】

【数１２】

$$f(x)=\sum_{i}^{n} w_i x_i + b \quad \cdots\cdots (12)$$

※式

【０２３１】

【数１３】

$$y=(y1, y2, \cdots\cdots, yn) \quad \cdots\cdots (13)$$

★【０２３４】

【数１４】

$$y_i(w' x_i + b) \geq 1 \quad \cdots\cdots (14)$$

☆ち、次式

【０２３７】

【数１５】

◆【０２３９】

【数１６】

$$\frac{\partial L}{\partial b}=\frac{\partial L}{\partial w}=0 \quad \cdots\cdots (16)$$

＊別を

【０２４２】

【数１７】

$$\max \sum ai-\frac{1}{}\sum aiaiyiyi' xj$$

制約条件：$ai \geqq 0, \sum aiyi=0$ …… （１７）

※制約条件を次式

【０２４５】

【数１８】

$$y_i (w' x_i + b) \geqq 1 - \xi_i \quad \cdots\cdots (18)$$

37

【０２４６】のように変更する。

【０２４７】最適化については、次式

$$\frac{1}{2}||w||^2 + C\sum\xi_i \qquad \cdots\cdots (19)$$

【０２４９】の目的関数を最小化する。

【０２５０】この（１９）式において、Ｃは、制約条件をどこまで緩めるかを指定する係数であり、実験的に値を決定する必要がある。

$$\max \sum ai - \frac{1}{2} \sum aiaiyiyi' xj$$

$$\text{制約条件}：0 \leqq ai \leqq C, \sum aiyi=0 \qquad \cdots\cdots (20)$$

【０２５３】のように変更される。

【０２５４】しかし、この（２０）式のままでは、非線型の問題を解くことはできない。そこで、本実施の形態では、カーネル関数Ｋ（ｘ、x^3）を導入して、一旦、高次元の空間に写像して（カーネル・トリック）、その

$$K(x,y)= \Phi(x)' \Phi(x')$$

【０２５７】のように表される。

【０２５８】また、（１２）式に示した識別関数も、次式

$$f(\Phi(x))=w' \Phi(x) + b$$

$$= \sum aiyiK(x,xi) + b \qquad \cdots\cdots (22)$$

【０２６０】のように表すことができる。

【０２６１】また学習に関しても、次式

$$\max \sum ai - \frac{1}{2}\sum aiaiyiyixjK(Xi,xj)$$

$$\text{制約条件}：0 \leqq ai \leqq C, \sum aiyi=0 \qquad \cdots\cdots (23)$$

【０２６３】に示す２次計画問題としてとらえることができる。

【０２６４】カーネルとしては、次式

$$K(x,x')=\exp\left(-\frac{|x-x'|^2}{\sigma^2}\right) \qquad \cdots\cdots (24)$$

【０２６６】に示すガウシアン・カーネル（ＲＢＦ（Radius Basic Function））などを用いることができる。

【０２６７】なお、ガボア・フィルタリングに関しては、認識タスクに応じてフィルタの種類を変更するようにしても良い。

【０２６８】低周波でのフィルタリングでは、フィルタリング後のイメージすべてをベクトルとして持っているのは冗長である。そこで、ダウンサンプリングして、ベクトルの次元を落とすようにしても良い。ダウンサンプリングされた２４種類のベクトルを一列に並べ、長いベクトルにする。

【０２６９】また本実施の形態において顔パターンの認識に適用されるＳＶＭは、特徴空間を２分する識別器なので、「人Ａ」か「人Ａでない」かを判別するように学習する。そのため、データベースの画像中から、まず人Ａの顔画像を集め、ガボア・フィルタリング後のベクトルに「人Ａでない」というラベルを貼る。一般に、集める顔画像の数は、特徴空間の次元より多い方がよい。１０人の

38

＊【０２４８】

【数１９】

※【０２５１】ラグランジュ定数ａに関する問題は次式

【０２５２】

【数２０】

★空間で線形分離することにしている。したがって、元の空間では非線型分離していることと同等となる。

【０２５５】カーネル関数は、ある写像Φを用いて次式

【０２５６】

【数２１】

$$\cdots\cdots (21)$$

☆【０２５９】

【数２２】

◆【０２６２】

【数２３】

＊【０２６５】

【数２４】

顔を認識したい場合は、同様に、「人Ｂである」、「人Ｂでない」…のように、それぞれの人に対して１つの識別器を構成する。

【０２７０】このような学習により、例えば、「人Ａ」と「人Ａでない」を分けるサポート・ベクタが求まる。ＳＶＭは、特徴空間を２つに仕切る識別器であり、新しい顔画像が入力されてきたときに、やはりガボア・フィルタリングのベクトルが、求めたサポート・ベクタが構成する境界面のどちら側にあるかで認識結果を出力する。そして、境界に対して、「人Ａ」の領域にあれば「人Ａ」と認識することができる。また、「人Ａではない」領域であれば「人Ａでない」と認識される。

【０２７１】ＣＣＤカメラ５０からの画像信号Ｓ１Ａに基づく画像から顔の部分として切り取られる領域は一定ではない。このため特徴空間で認識したいカテゴリとは離れた点に投影される可能性がある。従って、目や鼻、口といった特徴をもつパーツに推定してアフィン変換によりモーフィングすることにより認識率が向上する可能性がある。

【０２７２】また認識性能を上げるために、ブートストラップ手法を採用することができる。学習に用いる画像とは別に画像を撮影して、ブートストラップに用いる。これは、学習した識別器が誤った認識結果を出したときに、その入力画像を学習セットに投入して学習し直すことを意味する。

【０２７３】また認識性能を上げるために、認識結果の時間変化を見る方法もある。最も簡単な方法では、１０回中８回「人Ａ」と認識されたら「人Ａ」と認識するなどである。他に、カルマン・フィルタを用いた予測法なども提案されている。

【０２７４】（５）本実施の形態の動作及び効果

以上の構成において、このロボット１では、新規な人との対話を通してその人の名前を取得し、当該名前を、マイクロホン５１やＣＣＤカメラ５０の出力に基づいて検出したその人の声の音響的特徴及び顔の形態的特徴の各データと関連付けて記憶すると共に、これら記憶した各種データに基づいて、名前を取得していないさらに新規な人の登場を認識し、その新規な人の名前や声の音響的特徴及び顔の形態的特徴を上述と同様にして取得し記憶するようにして、人の名前を学習する。

【０２７５】従って、このロボット１は、音声コマンドの入力やタッチセンサの押圧操作等のユーザからの明示的な指示による名前登録を必要とすることなく、人間が普段行うように、通常の人との対話を通して新規な人物や物体等の名前を自然に学習することができる。

【０２７６】以上の構成によれば、新規な人との対話を通してその人の名前を取得し、当該名前を、マイクロホン５１やＣＣＤカメラ５０の出力に基づいて検出したその人の声の音響的特徴及び顔の形態的特徴の各データと関連付けて記憶すると共に、これら記憶した各データに基づいて、名前を取得していないさらに新規な人の登場を認識し、その新規な人の名前や声の音響的特徴及び顔の形態的特徴を上述と同様にして取得し記憶するようにして、人の名前を学習するようにしたことにより、通常の人との対話を通して新規な人物や物体等の名前を自然に学習し得るようにすることができ、かくしてエンターテインメント性を格段的に向上させ得るロボットを実現できる。

【０２７７】（６）他の実施の形態

なお上述の実施の形態においては、本発明を図１のように構成された２足歩行型のロボット１に適用するようにした場合について述べたが、本発明はこれに限らず、この他種々のロボット装置及びロボット装置以外のこの他種々の装置に広く適用することができる。

【０２７８】また上述の実施の形態においては、人間と対話するための機能を有し、当該対話を通して対象とする物体の名前を人間から取得する対話手段を、音声認識部６０、対話制御部６３及び音声合成部６４から構成することにより、人との音声対話により人の名前を取得するようにした場合について述べたが、本発明はこれに限らず、例えばキーボード入力等による文字対話により人の名前を取得するように対話手段を構成するようにしても良い。

【０２７９】さらに上述の実施の形態においては、名前学習の対象が人物である場合について述べたが、本発明はこれに限らず、人物に代えて又は人物に加えて以外のこの他種々の物体を名前学習の対象とするようにしても良い。

【０２８０】この場合において、上述の実施の形態においては、対象となる人物の声の音響的特徴及び顔の形態的特徴からその人物をそれぞれ認識し、これらの認識結果に基づいてその人物が新規な人物であるか否かを判別するようにした場合について述べたが、本発明はこれに限らず、これに代えて又はこれに加えて、これら以外の例えば体型やにおい等の生物学的に個体を識別可能な複数種類の各種特徴からその人物をそれぞれ認識し、これらの認識結果に基づいてその人物が新規な人であるか否かを判別するようにしても良い。また名前学習対象が人物以外の物体である場合には、色や形状、模様、大きさ等の物体を識別可能な複数種類の特徴からそれぞれその物体を認識し、これらの認識結果に基づいてその物体が新規な物体であるか否かを判別するようにしても良い。そしてこれらの場合には、それぞれ物体の異なる所定の特徴を検出すると共に、当該検出結果及び予め記憶している既知の物体の対応する特徴のデータに基づいて、当該対象とする物体を認識する複数の認識手段を設けるようにすれば良い。

【０２８１】さらに上述の実施の形態においては、既知の物体の名前及び当該物体に対する各認識手段（話者認識部６１及び顔認識部６２）の認識結果を関連付けた関

連付け情報を記憶する記憶手段をメモリにより構成するようにした場合について述べたが、本発明はこれに限らず、情報を記憶できるメモリ以外の例えばディスク状記録媒体等のこの他種々の記憶手段を広く適用することができる。

【０２８２】さらに上述の実施の形態においては、話者認識部６１及び顔認識部６２が対象とする人を認識する認識処理を１度しか行わないようにした場合について述べたが、本発明はこれに限らず、例えば認識不能（ＳＩＤ＝－１）であった場合にはもう１度認識処理を行うようにするようにしても良く、これ以外のときであっても複数回の認識処理を行うようにしても良い。このようにすることによって認識結果の精度を向上させることができる。

【０２８３】さらに上述の実施の形態においては、対話制御部６３が複数の認識手段（音声認識部６０、話者認識部６１、顔認識部６２）の認識結果の多数決により、その人が新規な人であるか否かを判断するようにした場合について述べたが、本発明はこれに限らず、多数決以外の手法によりこれら複数の認識手段の各認識結果に基づいてその人が新規な人であるか否かを判断するようにしても良い。

【０２８４】この場合において、例えば複数の認識手段の各認識結果に、その認識手段の認識性能に応じて重み付けをして、その重み付けした各認識結果に基づいて対象とする物体が新規なものであるか否かを判断する方法や、最も認識性能の高い認識手段と他の１つの認識手段の認識結果に基づき新規な人と判断できた場合には、他の認識手段の認識結果を利用しない方法等など種々の方法を広く適用することができる。

【０２８５】さらに上述の実施の形態においては、話者認識部６１や顔認識部６２が対象とする人を正しく認識できた場合にその話者認識部６１及び又は顔認識部６２に追加学習させることにより、統計的な安定によって認識精度を向上させるようにした場合について述べたが、本発明はこれに限らず、メモリ６５に格納される関連付け情報についても、同様に、何度も同じ組み合わせを覚えることによってその関連付け情報の信頼性を向上させ得るような機能を設けるようにしても良い。具体的には、このような機能の具現化方法として、例えば『電子情報通信学会論文誌,D-II,Vol.J82-D-II,No6,pp.1072-1081.』に記載されたニューラルネットを用いた方法を利用することができる。

【０２８６】

【発明の効果】以上のように本発明によれば、学習装置において、対話を通して対象とする物体の名前を取得する対話手段と、対象とする物体の複数の特徴のデータを検出し、当該検出結果及び既知の物体の対応する特徴のデータに基づいて、対象とする物体を認識する複数の認識手段と、既知の物体の名前に対する各認識手段の認識結果を関連付けた関連付け情報を記憶する記憶手段と、対話手段が取得した対象とする物体の名前、対象とする物体に対する各認識手段の認識結果、及び記憶手段が記憶する関連付け情報に基づいて、対象とする物体が新規な物体であるか否かを判断する判断手段と、判断手段が対象とする物体を新規な物体と判断したときに、当該対象とする物体に対応する複数の特徴のデータを各認識手段に記憶させると共に、当該対象とする物体についての関連付け情報を記憶手段に記憶させる制御手段とを設けるようにしたことにより、人間が普段行うように、通常の人との対話を通して新規な人物や物体等の名前を自然に学習することができ、かくしてエンターテイメント性を格段的に向上させ得る学習装置を実現できる。

【０２８７】また本発明によれば、学習方法において、対話を通して対象とする物体の名前を取得する対話ステップと、対象とする物体の複数の特徴のデータを検出し、当該検出結果及び既知の物体の対応する特徴のデータに基づいて、対象とする物体を認識する複数の認識ステップと、既知の物体の名前に対する各認識手段の認識結果を関連付けた関連付け情報を記憶する記憶ステップと、対話手段が取得した対象とする物体の名前、対象とする物体に対する各認識手段の認識結果、及び記憶手段が記憶する関連付け情報に基づいて、対象とする物体が新規な物体であるか否かを判断する判断ステップと、判断手段が対象とする物体を新規な物体と判断したときに、当該対象とする物体に対応する複数の特徴のデータを各認識手段に記憶させると共に、当該対象とする物体についての関連付け情報を記憶手段に記憶させる制御ステップとを設けるようにしたことにより、人間が普段行うように、通常の人との対話を通して新規な人物や物体等の名前を自然に学習することができ、かくしてエンターテイメント性を格段的に向上させ得る学習方法を実現できる。

【０２８８】さらに本発明によれば、ロボット装置において、対話を通して対象とする物体の名前を取得する対話手段と、対象とする物体の複数の特徴のデータを検出し、当該検出結果及び既知の物体の対応する特徴のデータに基づいて、対象とする物体を認識する複数の認識手段と、既知の物体の名前に対する各認識手段の認識結果を関連付けた関連付け情報を記憶する記憶手段と、対話手段が取得した対象とする物体の名前、対象とする物体に対する各認識手段の認識結果、及び記憶手段が記憶する関連付け情報に基づいて、対象とする物体が新規な物体であるか否かを判断する判断手段と、判断手段が対象とする物体を新規な物体と判断したときに、当該対象とする物体に対応する複数の特徴のデータを各認識手段に記憶させると共に、当該対象とする物体についての関連付け情報を記憶手段に記憶させる制御手段とを設けるようにしたことにより、人間が普段行うように、通常の人との対話を通して新規な人物や物体等の名前を自然に学

習することができ、かくしてエンターテイメント性を格段的に向上させ得るロボット装置を実現できる。

【０２８９】

【図面の簡単な説明】

【図１】本実施の形態によるロボットの外観構成を示す斜視図である。

【図２】本実施の形態によるロボットの外観構成を示す斜視図である。

【図３】本実施の形態によるロボットの外観構成の説明に供する略線図である。

【図４】本実施の形態によるロボットの内部構成の説明に供する略線図である。

【図５】本実施の形態によるロボットの内部構成の説明に供する略線図である。

【図６】名前学習機能に関するメイン制御部４０の処理の説明に供するブロック図である。

【図７】メモリにおけるＦＩＤ及びＳＩＤと名前との関連付けの説明に供する概念図である。

【図８】名前学習処理手順を示すフローチャートである。

【図９】名前学習処理手順を示すフローチャートである。

【図１０】名前学習処理時における対話例を示す略線図である。

【図１１】名前学習処理時における対話例を示す略線図である。

【図１２】ＦＩＤ及びＳＩＤと名前との新規登録の説明に供する概念図である。

【図１３】名前学習時における対話例を示す略線図である＊る。

【図１４】名前学習処理時における対話例を示す略線図である。

【図１５】音声認識部の構成を示すブロック図である。

【図１６】単語辞書の説明に供する概念図である。

【図１７】文法規則の説明に供する概念図である。

【図１８】特徴ベクトルバッファの記憶内容の説明に供する概念図である。

【図１９】スコアシートの説明に供する概念図である。

【図２０】音声認識処理手順を示すフローチャートである。

【図２１】未登録語処理手順を示すフローチャートである。

【図２２】クラスタ分割処理手順を示すフローチャートである。

【図２３】シミュレーション結果を示す概念図である。

【図２４】学習時における顔認識部の構成を示すブロック図である。

【図２５】認識時における顔認識部の構成を示すブロック図である。

【符号の説明】

１……ロボット、４０……メイン制御部、５０……ＣＣＤカメラ、５１……マイクロホン、５４……スピーカ、６０……音声認識部、６１……話者認識部、６２……顔認識部、６３……対話制御部、６４……音声合成部、６５……メモリ、Ｓ１Ａ……画像信号、Ｓ１Ｂ、Ｓ３……音声信号、Ｄ１、Ｄ２……文字列データ、ＲＴ１……名前学習処理手順。

【図7】

ＦＩＤ	ＳＩＤ	名前
1	2	フジタ
2	5	ヨシダ

図7　ＦＩＤ及びＳＩＤと名前との関連付け

【図１０】

R:フジタさんですよね。
H:はい、そうです。
R:今日はいい天気ですね。
H:そうですね。
・
・
・
・

図１０　対応例（１）

【図１１】

R:あ、フジタさんですよね。
H:いいえ違います。
R:あれ、名前を教えてください。
H:ヤマモトです。
R:私はロボットです。よろしくお願いします。
H:よろしくお願いします。
R:ヤマモトさん、今日はいい天気ですね。
H:そうですね。

図１１　対応例（２）

【図１２】

ＦＩＤ	ＳＩＤ	名前
1	2	フジタ
2	5	ヨシダ
4	6	ヤマモト

図１２　ＦＩＤ及びＳＩＤと名前の新規登録

【図１３】

R:あ、フジタさんですよね。
H:いいえ違います。
R:あれ、名前を教えてください。
H:ヨシダです。
R:ああヨシダさんですね。思いだしましたよ。今日はいい天気ですね。
H:そうですね。
R:前回はえーと、いつ会いましたっけ？
H:ええと、おとといじゃなかったかな？

図１３　対応例（３）

【図１】

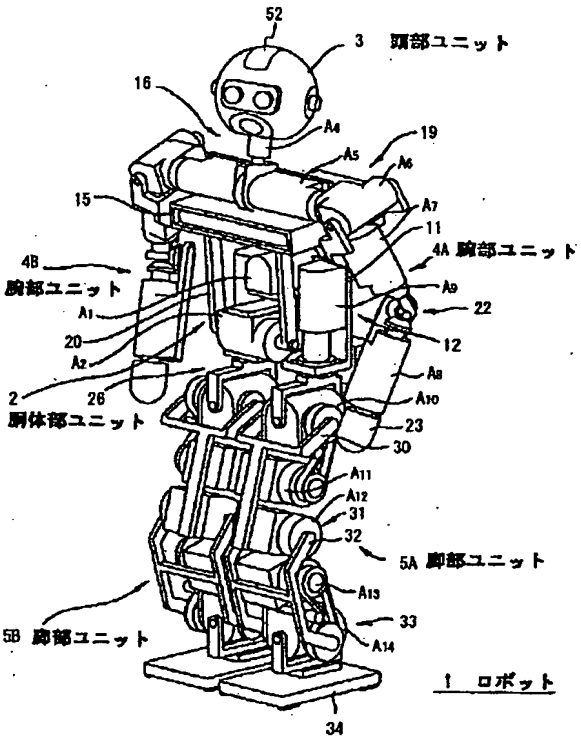

図１　本実施の形態におけるロボットの外観構成（１）

【図２】

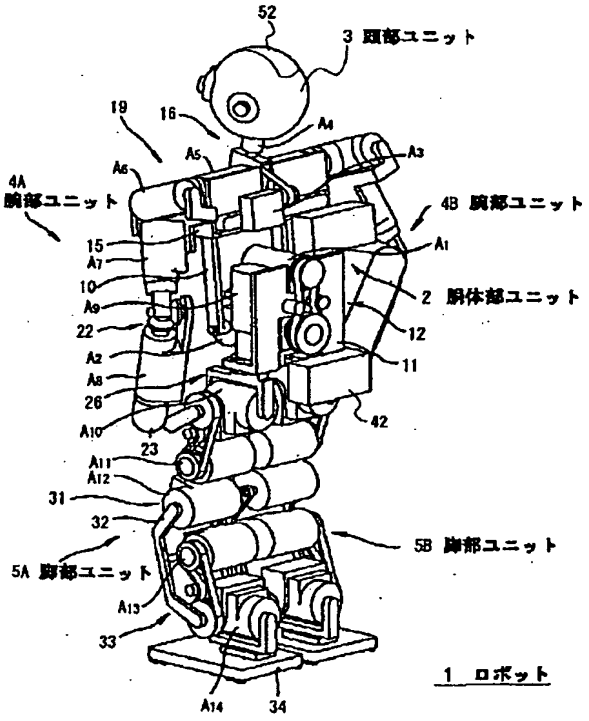

図２　本実施の形態によるロボットの外観構成（２）

【図５】

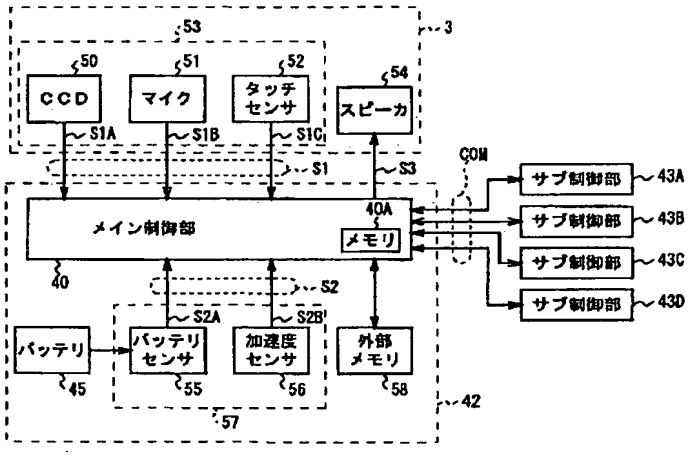

図５　ロボットの内部構成（２）

【図８】

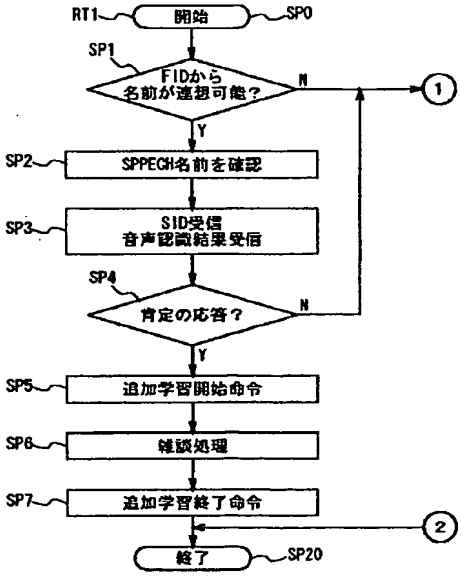

図８　名前学習処理手順（１）

【図３】

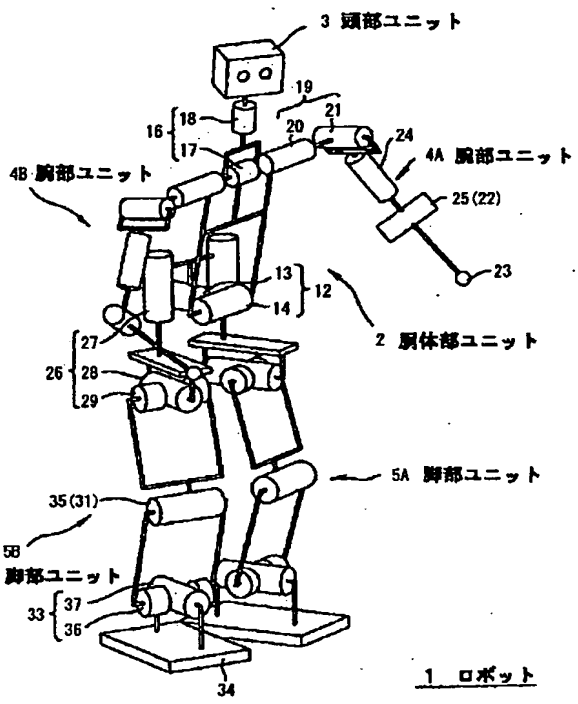

図３　本実施の形態によるロボットの外観構成（３）

【図４】

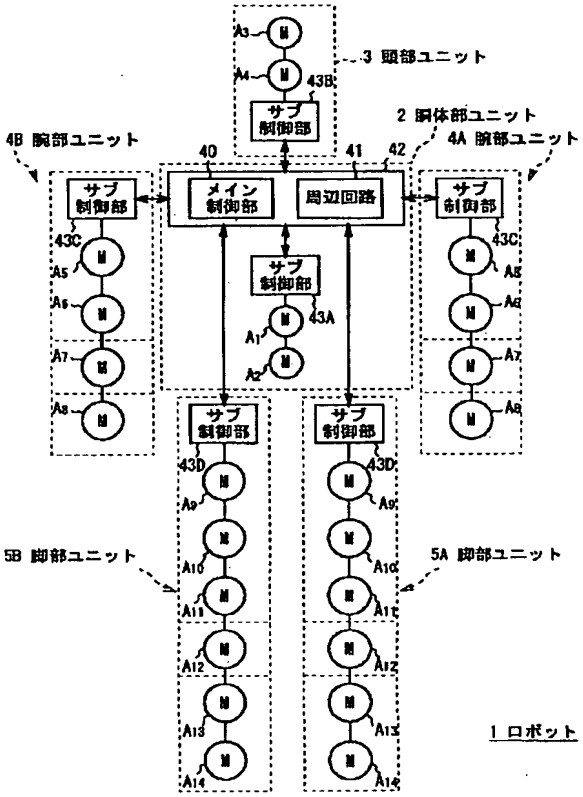

図４　ロボットの内部構成（１）

【図６】

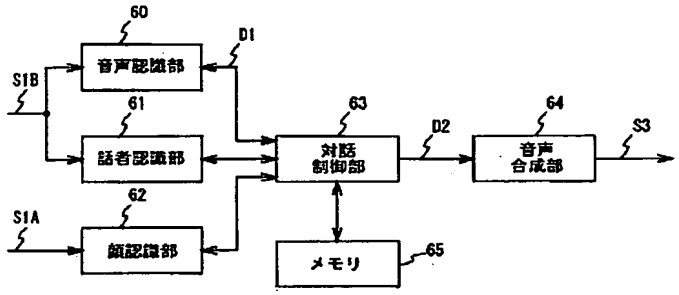

図６　メイン制御部の処理

【図９】

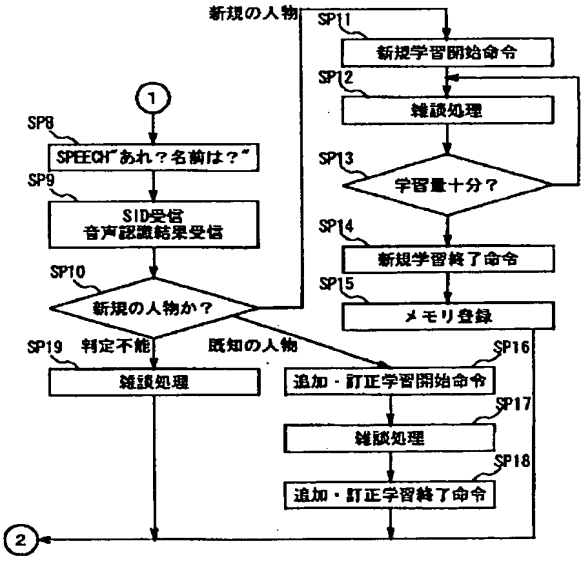

図９　名前学習処理手順（２）

【図１４】

(R:あ、フジタさんですよね。)
(H:いいえ違います。
R:あれ、名前を教えてください。
H:ヤマモトです。
R:ああそうですか。元気ですか？
H:ええ元気ですよ。

図１４　対応例（４）

【図１５】

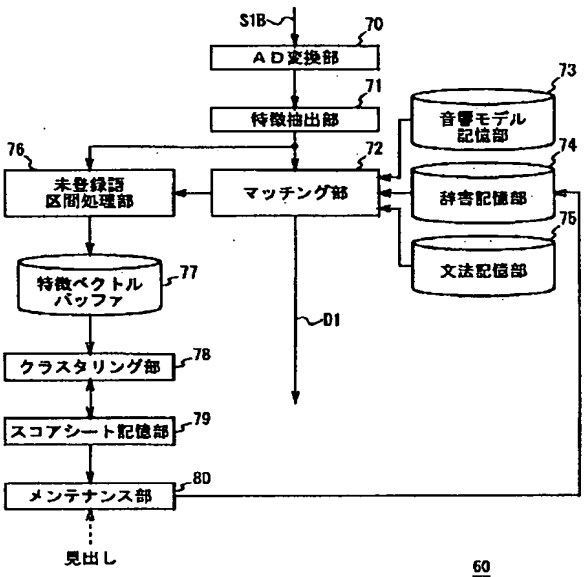

図１５　音声認識部の構成

【図１６】

見出し		音韻系列
boku	[僕]	boku
chigau	[違う]	chigau
doko	[どこ]	doko
genki	[元気]	geNki
iro	[色]	iro
janai	[じゃない]	janai
kirai	[嫌い]	kirai
kudasai	[ください]	kudasai

図１６　単語辞書

【図１７】

```
$col    =  [kono | sono] iro wa;
$this   =  kore [(ga | wa | mo)];
$neg    =  (Chagau | iie) [$sil];
$null   =  $sil;
$des    =  (desu | da) [yo] | yo;
$not    =  janai [yo];
$color1 =  $null | $neg | [$neg] $col | [$neg] $this;
$color2 =  [iro] (desu | janai | da) [yo];
$pat1   =  $color $garbage $color2;
```

図１７　文法規則

【図１８】

ID	音韻系列	特徴ベクトル系列
1	—	—
2	—	—
・	・	・
・	・	・
・	・	・
N	—	—
N+1	—	—

図１８　特徴ベクトルバッファ

【図１９】

ID	音韻系列	クラスタナンバ	代表メンバID	スコア(距離) 1	2	3	…	N	N+1
1	…	1	1	s(1,1)	s(1,2)	s(1,3)	…	s(1,N)	s(1,N+1)
2	…	2	2	s(2,1)	s(2,2)	s(2,3)	…	s(2,N)	s(2,N+2)
3	…	1	1	s(3,1)	s(3,2)	s(3,3)	…	s(3,N)	s(3,N+3)
・	・	・	・	・	・	・	・	・	・
・	・	・	・	・	・	・	・	・	・
・	・	・	・	・	・	・	・	・	・
				…					
N	…	1	1	s(N,1)	s(N,2)	s(N,3)	…	s(N,N)	s(N,N+1)
N+1	…	2	2	s(N+1,1)	s(N+1,2)	s(N+1,3)	…	s(N+1,N)	s(N+1,N+1)

図１９　スコアシート

【図２３】

音韻系列	発話単語
doroa:	風呂×1;
kuro	風呂×3;
Nfuro	風呂×20;
NhoNn	本×18;
hoNN	本×6;
NhoNde	本×10;
NhoNde:su	本×4;
ohoN	オレンジ×1;本×19;
hoNgdawasoNre:s:	本×2;
a:modori:	緑色×11;
ouidori:	緑色×10;
e:|midori:	緑色×3;
Nmidori:	緑色×5;
a:midori:|iroiresu	緑色×4;
Nro:ka	廊下×10;
Nro:kaNs	廊下×10;

図２３　シミュレーション結果

【図２０】

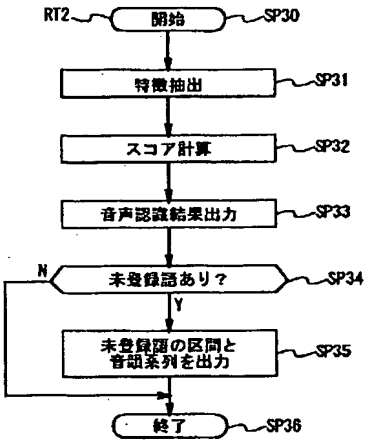

図２０　音声認識処理手順

【図２１】

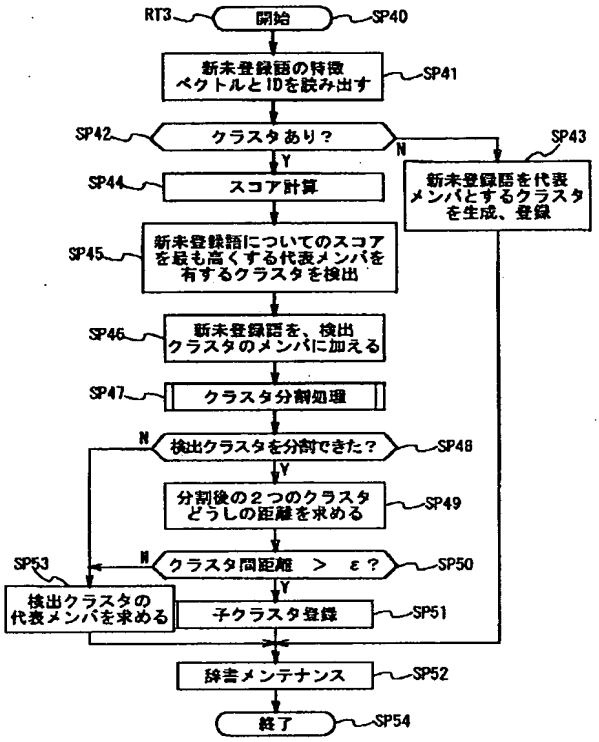

図２１　未登録語処理手順

【図２２】

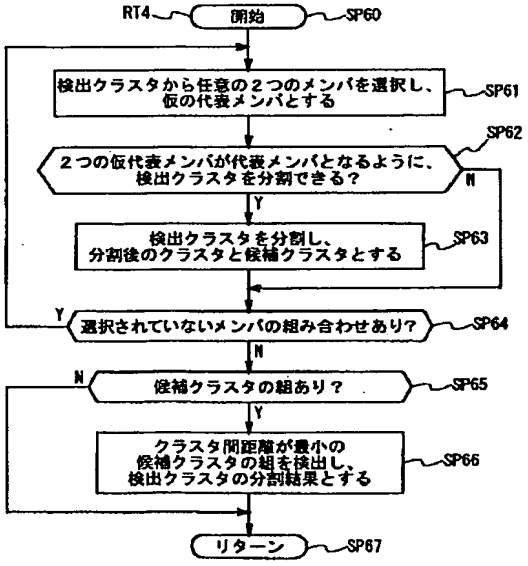

図２２　クラスタ分割処理手順

【図２４】

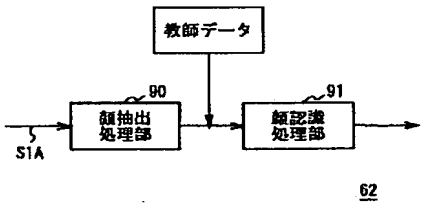

図２４　学習時における顔認識部の具体的構成

【図２５】

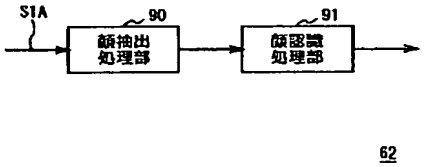

図２５　認識時における顔認識部の構成

フロントページの続き

(51)Int.Cl.[7]	識別記号	FI	
G10L	15/06	G10L 3/00	545F
	17/00		551H
			R

(72)発明者　浅野　康治
東京都品川区北品川6丁目7番35号ソニー株式会社内
(72)発明者　大久保　厚志
東京都品川区北品川6丁目7番35号ソニー株式会社内

(56)参考文献　特開2001－300148（ＪＰ，Ａ）
特開　平７－287695（ＪＰ，Ａ）
特開2001－228891（ＪＰ，Ａ）
特開2002－49424（ＪＰ，Ａ）
特開2002－202795（ＪＰ，Ａ）
特開2003－22131（ＪＰ，Ａ）
特開2003－186494（ＪＰ，Ａ）
特開2003－44080（ＪＰ，Ａ）
特開2002－219677（ＪＰ，Ａ）
Deb Roy, Integration of speech and vision using mutual information, Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, 米国, 2000年 6月 5日, Vol. 4, Pages 2369－2372
金, 岩橋, 知覚情報の統合に基づく言語音声単位の獲得アルゴリズム, 電子情報通信学会技術研究報告［思考と言語］, 日本, 2000年10月13日, ＴＬ2000－21, Ｐａｇｅｓ 9－16
下村, 青山, 藤田, 自立型エンタテイメントロボットと音声対話, 人工知能学会第36回言語・音声理解と対話処理研究会資料, 日本, 2002年11月 7日, Ｐａｇｅｓ 21－26

(58)調査した分野(Int.Cl.[7], ＤＢ名)
G10L 15/00 - 15/28
A63H 11/00
B25J 5/00
ＪＩＣＳＴファイル（ＪＯＩＳ）
ＩＥＥＥ Ｘｐｌｏｒｅ

CLAIMS

---

(57) [Claim(s)]

[Claim 1] The data of two or more descriptions of a dialogue means to acquire the identifier of the target body through a dialogue, and the body made into the above-mentioned object are detected. Two or more recognition means to recognize the body made into the above-mentioned object based on the data of the description with which the detection result concerned and a known body correspond, A storage means which associated the recognition result of each above-mentioned recognition means against the identifier of the above-mentioned known body to relate and to memorize information, The identifier of the body made into the above-mentioned object which the above-mentioned dialogue means acquired, the recognition result of each above-mentioned recognition means against the body made into the

above-mentioned object, And when the body which a decision means to judge whether the body made into the above-mentioned object is a new body, and the above-mentioned decision means make the above-mentioned object is judged to be a new body based on the above-mentioned correlation information which the above-mentioned storage means memorizes Study equipment characterized by having the control means which it relates [control means] about the body made into the object concerned, and makes the above-mentioned storage means memorize information while making each above-mentioned recognition means memorize the data of two or more above-mentioned descriptions corresponding to the body made into the object concerned. [Claim 2] The above-mentioned control means is study equipment according to claim 1 characterized by controlling an above-mentioned recognition means by which the body made into the object concerned has been recognized correctly to carry out additional study when the body which the above-mentioned decision means makes the above-mentioned object is judged to be the above-mentioned known body.

[Claim 3] The above-mentioned control means is study equipment according to claim 1 characterized by controlling an above-mentioned recognition means by which the body made into the object concerned has not been recognized correctly to carry out correction study when the body which the above-mentioned

decision means makes the above-mentioned object is judged to be the above-mentioned known body.

[Claim 4] The above-mentioned decision means is study equipment according to claim 1 which carries out [judging whether the body which makes into the above-mentioned object is a new body by the majority of the identifier of the body made into the above-mentioned object which the above-mentioned dialogue means acquired, and the recognition result of each above-mentioned recognition means against the body concerned, referring to the above-mentioned correlation information which the above-mentioned storage means memorizes, and] as the description.

[Claim 5] The above-mentioned control means is study equipment according to claim 1 characterized by controlling the above-mentioned dialogue means to extend a dialogue if needed.

[Claim 6] The data of two or more descriptions of the dialogue step which acquires the identifier of the target body through a dialogue, and the body made into the above-mentioned object are detected. Two or more recognition steps which recognize the body made into the above-mentioned object based on the data of the description with which the detection result concerned and a known body correspond, The storage step which associated the recognition result of each above-mentioned recognition means against the identifier of the

above-mentioned known body and which relates and memorizes information, The identifier of the body made into the above-mentioned object which the above-mentioned dialogue means acquired, the recognition result of each above-mentioned recognition means against the body made into the above-mentioned object, And when the decision step which judges whether the body made into the above-mentioned object is a new body, and the body which the above-mentioned decision means makes the above-mentioned object are judged to be new bodies based on the above-mentioned correlation information which the above-mentioned storage means memorizes The study approach characterized by having the control step which it relates [step] about the body made into the object concerned, and makes the above-mentioned storage means memorize information while making each above-mentioned recognition means memorize the data of two or more above-mentioned descriptions corresponding to the body made into the object concerned. [Claim 7] The study approach according to claim 6 characterized by carrying out additional study about the above-mentioned description which has recognized correctly the body made into the object concerned when the body made into the above-mentioned object is judged to be the above-mentioned known body at the above-mentioned control step.

[Claim 8] The study approach according to claim 6 characterized by carrying out

correction study about the above-mentioned description which has not recognized correctly the body made into the object concerned when the body made into the above-mentioned object is judged to be the above-mentioned known body at the above-mentioned control step.

[Claim 9] The study approach according to claim 6 that the body made into the above-mentioned object is characterized by judging whether it is a new body at the above-mentioned decision step by the majority of each recognition result based on the identifier of the body made into the acquired above-mentioned object, and each above-mentioned description of the body concerned, respectively while referring to the above-mentioned correlation information.

[Claim 10] The study approach according to claim 6 characterized by extending the dialogue concerned if needed at the above-mentioned dialogue step.

[Claim 11] The data of two or more descriptions of a dialogue means to acquire the identifier of the target body through a dialogue, and the body made into the above-mentioned object are detected. Two or more recognition means to recognize the body made into the above-mentioned object based on the data of the description with which the detection result concerned and a known body correspond, A storage means which associated the recognition result of each above-mentioned recognition means against the identifier of the above-mentioned known body to relate and to memorize information, The

identifier of the body made into the above-mentioned object which the above-mentioned dialogue means acquired, the recognition result of each above-mentioned recognition means against the body made into the above-mentioned object, And when the body which a decision means to judge whether the body made into the above-mentioned object is a new body, and the above-mentioned decision means make the above-mentioned object is judged to be a new body based on the above-mentioned correlation information which the above-mentioned storage means memorizes Robot equipment characterized by having the control means which it relates [control means] about the body made into the object concerned, and makes the above-mentioned storage means memorize information while making each above-mentioned recognition means memorize the data of two or more above-mentioned descriptions corresponding to the body made into the object concerned. [Claim 12] The above-mentioned control means is robot equipment according to claim 11 characterized by controlling an above-mentioned recognition means by which the body made into the object concerned has been recognized correctly to carry out additional study when the body which the above-mentioned decision means makes the above-mentioned object is judged to be the above-mentioned known body.

[Claim 13] The above-mentioned control means is robot equipment according to

claim 11 characterized by controlling an above-mentioned recognition means by which the body made into the object concerned has not been recognized correctly to carry out correction study when the body which the above-mentioned decision means makes the above-mentioned object is judged to be the above-mentioned known body.

[Claim 14] the body which carries out as the above-mentioned object by the majority of the identifier of the body which makes into the above-mentioned object which the above-mentioned dialogue means acquired while the above-mentioned decision means refers to the above-mentioned correlation information which the above-mentioned storage means memorizes, and the recognition result of each above-mentioned recognition means against the body concerned -- the above -- the robot equipment according to claim 11 characterized by to judge whether it is a new body.

[Claim 15] The above-mentioned control means is robot equipment according to claim 11 characterized by controlling the above-mentioned dialogue means to extend a dialogue if needed.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention is applied to an entertainment robot, concerning robot equipment, and is suitable for study equipment and the study approach list.

[0002]

[Description of the Prior Art] In recent years, many entertainment robots for ordinary homes are commercialized. And various external sensors, such as a CCD (ChargeCoupled Device) camera and a microphone, are carried, an outside circumference is recognized based on the output of these external sensor, and there are some entertainment robots made as [act / based on a recognition result / it / autonomously].

[0003]

[Problem(s) to be Solved by the Invention] By the way, it sets to this entertainment robot and is a new body (a person also contains.). Hereafter, it is the same. If an identifier is matched with the body and it can be memorized, communication with a user can be made smoother and it will be thought that it can make it possible to correspond also to the various instructions for bodies other than the body with which the identifier was registered beforehand "kick a ball" from a user, flexibly. In addition, it shall express matching an objective

identifier with the body and memorizing it to below as mentioned above, as "an identifier is learned", and such a function shall be called an "identifier learning function."

[0004] Moreover, if an entertainment robot can make it possible to learn the identifier of a new body through a dialogue with the usual man so that it may face carrying such an identifier learning function in an entertainment robot and human being may carry out usually, it will think from the natural gender and it will be thought that it is the most desirable and the entertainment nature as an entertainment robot may be raised further.

[0005] However, with the conventional technique, there is a problem with difficult making an entertainment robot judge when the new body which should learn an identifier has appeared at hand.

[0006] For this reason, in the former, after carrying out press actuation of the specific touch sensor which the user gave the explicit voice command or was arranged by the robot and changing a mode of operation into register mode, many technique of performing objective recognition and registration of that identifier is used. However, when considering the natural interaction of a user and an entertainment robot, the identifier registration by such explicit directions had a truly unnatural problem.

[0007] This invention was made in consideration of the above point, and tends to

propose robot equipment in the study equipment and the study approach list which may raise entertainment nature on a target markedly.

[0008]

[Means for Solving the Problem] In order to solve this technical problem, it sets to this invention. A dialogue means to acquire the identifier of the target body through a dialogue in study equipment, Two or more recognition means to recognize the target body based on the data of the description with which the data of two or more descriptions of the target body are detected, and the detection result concerned and a known body correspond, A storage means which associated the recognition result of each recognition means against the identifier of a known body to relate and to memorize information, A decision means to judge whether the target body is a new body based on the identifier of the body made into the object which the dialogue means acquired, the recognition result of each recognition means against the target body, and the correlation information that a storage means memorizes, When a decision means judged the target body to be a new body, while making each recognition means memorize the data of two or more descriptions corresponding to the body made into the object concerned, the control means which it relates [control means] about the body made into the object concerned, and makes a storage means memorize information was established.

[0009] As a result, without needing the identifier registration by the explicit directions from users, such as an input of a voice command, and press actuation of a touch sensor, this study equipment can learn the identifier of a new person, a body, etc. automatically through a dialogue with the usual man so that human being may carry out usually.

[0010] Moreover, the dialogue step which acquires the identifier of the target body through a dialogue in the study approach in this invention, Two or more recognition steps which recognize the target body based on the data of the description with which the data of two or more descriptions of the target body are detected, and the detection result concerned and a known body correspond, The storage step which associated the recognition result of each recognition means against the identifier of a known body and which relates and memorizes information, The decision step which judges whether the target body is a new body based on the identifier of the body made into the object which the dialogue means acquired, the recognition result of each recognition means against the target body, and the correlation information that a storage means memorizes, When a decision means judged the target body to be a new body, while making each recognition means memorize the data of two or more descriptions corresponding to the body made into the object concerned, the control step which it relates [step] about the body made into the object concerned, and

makes a storage means memorize information was prepared.

[0011] Consequently, according to this study approach, without needing the identifier registration by the explicit directions from users, such as an input of a voice command, and press actuation of a touch sensor, the identifier of a new person, a body, etc. can be automatically learned through a dialogue with the usual man so that human being may carry out usually.

[0012] A dialogue means to acquire the identifier of the target body through a dialogue in robot equipment in this invention furthermore, Two or more recognition means to recognize the target body based on the data of the description with which the data of two or more descriptions of the target body are detected, and the detection result concerned and a known body correspond, A storage means which associated the recognition result of each recognition means against the identifier of a known body to relate and to memorize information, A decision means to judge whether the target body is a new body based on the identifier of the body made into the object which the dialogue means acquired, the recognition result of each recognition means against the target body, and the correlation information that a storage means memorizes, When a decision means judged the target body to be a new body, while making each recognition means memorize the data of two or more descriptions corresponding to the body made into the object concerned, the control means

which it relates [control means] about the body made into the object concerned, and makes a storage means memorize information was established.

[0013] Consequently, without needing the identifier registration by the explicit directions from users, such as an input of a voice command, and press actuation of a touch sensor, this robot equipment can learn the identifier of a new person, a body, etc. automatically through a dialogue with the usual man so that human being may carry out usually.

[0014]

[Embodiment of the Invention] About a drawing, the gestalt of 1 operation of this invention is explained in full detail below.

[0015] (1) While 1 shows the robot of the 2-pair-of-shoes walk mold by the gestalt of this operation as a whole and the head unit 3 is arranged in the upper part of the idiosoma unit 2 in a robot's block diagram 1 and drawing 2 by the gestalt of this operation It is constituted by arranging the arm units 4A and 4B of the respectively same configuration as up right and left of the idiosoma unit 2 concerned, respectively, and attaching the leg units 5A and 5B of the respectively same configuration as the lower left right of the idiosoma unit 2 in a predetermined location, respectively.

[0016] It is constituted when the waist base 11 which forms the frame 10 which forms the truncus upper part, and a lower trunk in the idiosoma unit 2 connects

through the waist joint device 12. It is made as [rotate / around the roll axes 13 which are shown in drawing 3 and which intersect perpendicularly, and a pitching axis 14 /, respectively / the truncus upper part / independently] by driving each actuators A1 and A2 of the waist joint device 12 fixed to the waist base 11 of a lower trunk, respectively.

[0017] Moreover, the head unit 3 is attached in the top-face center section of the shoulder base 15 fixed to the upper limit of a frame 10 through the neck joint device 16, and is made as [make / it / to rotate independently, respectively around the pitching axis 17 which is shown in drawing 3 and which intersects perpendicularly, and a yawing axis 18] by driving each actuator A3 of the neck joint device 16 concerned, and A4, respectively.

[0018] Furthermore, each arm units 4A and 4B are attached in right and left of the shoulder base 15 through the shoulder-joint device 19, respectively, and are made as [make / it / to rotate independently, respectively around each corresponding actuator A5 of the shoulder-joint device 19, the pitching axis 20 which is shown in drawing 3 by driving A6, respectively and which intersects perpendicularly, and roll axes 21].

[0019] In this case, the actuator A8 which forms the forearm section in the output shaft of the actuator A7 which forms the overarm section, respectively through the elbow-joint device 22 is connected, and each arm units 4A and 4B are

constituted by attaching a hand part 23 at the tip of the forearm section concerned.

[0020] And in each arm units 4A and 4B, it is made by driving an actuator A7 as [make / it can be made to be able to rotate around the yawing axis 24 which shows the forearm section to drawing 3 , and / it / to rotate, respectively around the pitching axis 25 which shows the forearm section to drawing 3 by driving an actuator A8].

[0021] on the other hand, each actuator of the hip joint device 26 which is attached in the waist base 11 of a lower trunk through the hip joint device 26, respectively, and corresponds in each leg units 5A and 5B, respectively -- A9-A11 -- it is made by driving, respectively as [make / it / to rotate independently, respectively around the yawing axis 27 which intersects perpendicularly mutually and roll axes 28 which are shown in drawing 3 , and a pitching axis 29].

[0022] In this case, each leg units 5A and 5B are constituted by connecting a foot 34 with the lower limit of the frame 32 concerned through the ankle joint device 33 while the frame 32 which forms the leg section in the lower limit of the frame 30 which forms a femoral region, respectively through the knee-joint device 31 is connected.

[0023] This sets to each leg units 5A and 5B. By driving the actuator A12 which

forms the knee-joint device 31 By being able to make it rotate around the pitching axis 35 which shows the leg section to drawing 3 , and driving the actuators A13 and A14 of the ankle joint device 33, respectively It is made as [make / it / to rotate independently, respectively around the pitching axis 36 which shows a foot 34 to drawing 3 and which intersects perpendicularly, and roll axes 37].

[0024] On the other hand, as shown in drawing 4 , the Maine control section 40 which manages the motion control of the robot 1 whole concerned, the circumference circuits 41, such as a power circuit and a communication circuit, and the control unit 42 with which a box comes to contain a dc-battery 45 (drawing 5) etc. are arranged in the tooth-back side of the waist base 11 which forms the lower trunk of the idiosoma unit 2.

[0025] And it connects with each sub control sections 43A-43D arranged, respectively in each configuration unit (the idiosoma unit 2, the head unit 3, each arm units 4A and 4B, and each leg units 5A and 5B), and this control unit 42 is made as [communicate / required supply voltage can be supplied or / with these sub control sections 43A-43D / it / to these sub control sections 43A-43D,].

[0026] Moreover, it connects with each actuators A1-A14 in the configuration unit which corresponds, respectively, and each sub control sections 43A-43D are made as [drive / it / in the condition of having been specified based on the

various control command to which each actuators A1-A14 in the configuration unit concerned are given from the Maine control section 40].

[0027] The external sensor section 53 which furthermore becomes the head unit 3 from a microphone 51, a touch sensor 52, etc. which function as the CCD (Charge Coupled Device) camera 50 which functions as this robot's 1 "eyes", and a "lug" as shown in drawing 5 , The loudspeaker 54 which functions as "opening" is arranged in a predetermined location, respectively, and the internal sensor section 57 which consists of a dc-battery sensor 55, an acceleration sensor 56, etc. is arranged in the control unit 42.

[0028] And while CCD camera 50 of the external sensor section 53 picturizes a surrounding situation and obtained picture signal S1A is sent out to the Maine control section, a microphone 51 collects various instruction voice given as voice input from a user, such as "walk", "lie down", or "pursue a ball", and is made as [send / to the Maine control section 40 / sound signal S1B obtained in this way].

[0029] Moreover, in drawing 1 and drawing 2 , the touch sensor 52 is formed in the upper part of the head unit 3 so that clearly, it detects the pressure received by "it strokes" and the physical influence of "striking" from a user, and sends out a detection result to the Maine control section 40 as pressure detecting-signal S1C.

[0030] Furthermore, while the dc-battery sensor 55 of the internal sensor section

57 detects the energy residue of a dc-battery 45 a predetermined period and sends out a detection result to the Maine control section 40 as dc-battery residue detecting-signal S2A, an acceleration sensor 56 detects the acceleration of 3 shaft orientations (a x axis, y-axis, and z-axis) a predetermined period, and it sends out a detection result to the Maine control section 40 as an acceleration detecting-signal S2B.

[0031] Picture signal S1A to which the Maine control-section section 40 is supplied, respectively from CCD camera 50, the microphone 51, and touch sensor 52 grade of the external sensor section 53, Sound signal S1B, pressure detecting-signal S1C (these are hereafter called collectively the external sensor signal S1), etc., Dc-battery residue detecting-signal S2A, acceleration detecting-signal S2B, etc. which are supplied, respectively from the dc-battery sensor 55, an acceleration sensor, etc. of the internal sensor section 57 It is based on (these are hereafter called collectively the internal sensor signal S2), and a robot's 1 perimeter and an internal situation, the existence of the command from a user and the influence from a user, etc. are judged.

[0032] And the Maine control section 40 opts for the action which continues based on this decision result, the control program beforehand stored in internal-memory 40A, and the various control parameters stored in the external memory 58 with which it is then loaded, and sends out the control command

based on a decision result to the corresponding sub control sections 43A-43D. Consequently, based on this control command, the actuators A1-A14 corresponding to the basis of control of those sub control sections 43A-43D drive, will make the head unit 3 rock vertically and horizontally in this way, the arm units 4A and 4B will be raised upwards, or action of walking will be discovered by the robot 1.

[0033] Moreover, the Maine control section 40 blinks this by making the voice based on the sound signal S3 concerned output outside, or outputting a driving signal to LED prepared in the predetermined location of the head unit 3 which functions as a "eye" on appearance by giving the predetermined sound signal S3 to a loudspeaker 54 if needed in this case.

[0034] Thus, in this robot 1, it is made as [act / based on the situation of a perimeter and the interior, the existence of the command from a user, and influence, etc. / it / autonomously].

[0035] (2) processing of the Maine control section 40 about an identifier learning function -- explain the identifier learning function carried in this robot 1 next.

[0036] To this robot 1, that man's identifier is acquired through a dialogue with people. The identifier concerned Relate with each data of the acoustical description of the voice of the man who detected based on the output of a microphone 51 or CCD camera 50, and the gestalt-description of a face, and

while memorizing Based on each these-memorized data, recognize an appearance of the new person who does not acquire the identifier, and acquire the new person's identifier, the acoustical description of voice, and the gestalt-description of a face like ****, and they are memorized. People's identifier is matched with the man and the identifier learning function which carries out acquisition (this is hereafter called study of identifier) study is carried. In addition, "a known man", and a call and those who have finished memorizing shall be called "a new person" for the person who matches with the acoustical description of the man's voice, and the gestalt-description of a face below, and finished memorizing an identifier.

[0037] And this identifier learning function is realized by the various processings in the Maine control section 40.

[0038] If the contents of processing of the Maine control section 40 about this identifier learning function are classified here functionally, as shown in drawing 6 The speech recognition section 60 which recognizes the language which people uttered, and the speaker-recognition section 61 which identifies and recognizes the man based on the detected acoustical description concerned while detecting the acoustical description of people's voice, The face recognition section 62 which identifies and recognizes the man based on the detected gestalt-description concerned while detecting the gestalt-description of people's

face, The dialogue control section 63 which manages the various control for identifier study of a new person including dialogue control with people, and the storage management of a known man's identifier, the acoustical description of voice, and the gestalt-description of a face, It can divide into the speech synthesis section 64 which generates the sound signal S3 for [various] a dialogue on the basis of control of the dialogue control section 63, and is sent out to a loudspeaker 54 (drawing 5).

[0039] In this case, in the speech recognition section 60, by performing predetermined speech recognition processing based on sound signal S1B from a microphone 51 (drawing 5), it has the function to recognize the language contained in the sound signal S1B concerned per word, and is made as [send / to the dialogue control section 63 / by using these recognized words as character-string data D1].

[0040] Moreover, the speaker-recognition section 61 has the function detected by predetermined signal processing using the approach indicated by "Segregation of Speakers for Recognition and Speaker Identification (CH2977-7/91/0000~0873 S1.00 1991 IEEE)" in the acoustical description of people's voice contained in sound signal S1B given from a microphone 51.

[0041] And the speaker-recognition section 61 carries out a sequential comparison with the data of the acoustical description of all the known men that

have usually sometimes memorized the data of this detected acoustical description then. When the acoustical description then detected is in agreement with the acoustical description of the man of either known, with the acoustical description of the known man concerned, and the matched acoustical description concerned The identifier of a proper While notifying (this is hereafter called SID) to the dialogue control section 63, when the detected acoustical description is in agreement with neither of the known men's acoustical descriptions, it is made as [notify / to the dialogue control section 63 / SID (= -1) which means recognition impossible].

[0042] Moreover, when the dialogue control section 63 judges that he is a new person, the speaker-recognition section 61 detects the acoustical description of that man's voice in the meantime based on the initiation instruction of new study and study termination instruction which are given from the dialogue control section 63 concerned, and it is made as [notify / to the dialogue control section 63 / this SID] while matching the data of the detected acoustical description concerned with SID of a new proper and memorizing them.

[0043] In addition, the speaker-recognition section 61 is made as [carry / according to an initiation instruction and a termination instruction of the additional study from the dialogue control section 63, or correction study / the additional study which collects the data of the acoustical description the man's

voice additionally, and the correction study which corrects the data of the acoustical description the man's voice so that the man can be recognized correctly].

[0044] In the face recognition section 62, picture signal S1A given from CCD camera 50 (drawing 5) is monitored continuously, and it has the function in which predetermined signal processing detects the gestalt-description of people's face contained in the image based on the picture signal S1A concerned.

[0045] And the face recognition section 62 carries out a sequential comparison with the data of the gestalt-description of the face of all the known men that have usually sometimes memorized the data of this detected gestalt-description then. With the gestalt-description concerned matched with the gestalt-description of the known man concerned when the gestalt-description then detected was in agreement with the gestalt-description of the face of the man of either known, the identifier of a proper While notifying (this is hereafter called FID) to a dialogue control section, when the detected gestalt-description is not in agreement with the gestalt-description of which known man's face, it is made as [notify / to a dialogue control section / FID (= -1) which means recognition impossible].

[0046] Moreover, it is based on the study initiation instruction and study termination instruction which are given from the dialogue control section 63

concerned when it judges that the face recognition section 62 is a person with the new dialogue control section 63. The gestalt-description of people's face contained in the meantime in the image based on picture signal S1A from CCD camera 50 is detected, and while matching the data of the detected gestalt-description concerned with FID of a new proper and memorizing them, it is made as [notify / to the dialogue control section 63 / this FID].

[0047] In addition, the face recognition section 62 is made as [carry / according to an initiation instruction and a termination instruction of the additional study from the dialogue control section 63, or correction study / the additional study which collects the data of the gestalt-description people's face additionally, and the correction study which corrects the data of the gestalt-description people's face so that the man can be recognized correctly].

[0048] The speech synthesis section 64 has the function to change into a sound signal S3 the character-string data D2 given from the dialogue control section 63, and is made as [send / to a loudspeaker 54 (drawing 5) / the sound signal S3 acquired in this way]. It is made as [make / by this / the voice based on this sound signal S3 / to output from a loudspeaker 54].

[0049] In the dialogue control section 63, as shown in drawing 7 , it has the memory 65 (drawing 6) which associates and memorizes SID matched with the data of the acoustical description of the voice of a known man's identifier, and its

man whom the speaker-recognition section 61 has memorized, and FID matched with the data of the gestalt-description the man's face which the face recognition section 62 has memorized.

[0050] The dialogue control section 63 and by giving the character-string data D2 predetermined to predetermined timing to the speech synthesis section 64 While making the voice for talking, and asking an identifier to a partner's man, or checking an identifier etc. output from a loudspeaker 54 The recognition result of the face recognition section [as opposed to that man to each recognition result list of the speech recognition section 60 based on a response of that man at this time etc., and the speaker-recognition section 61] 62, It is made as [judge / the man / whether you are a new person / based on an above-mentioned known man's identifier stored in memory 65, and the information on correlation of SID and FID].

[0051] and when it judges that the dialogue control section 63 is a person with the new man By giving an initiation instruction and a termination instruction of new study to the speaker-recognition section 61 and the face recognition section 62 While making these speaker-recognitions section 61 and the face recognition section 62 collect and memorize the data of the acoustical description of the new person's voice, or the gestalt-description of a face SID and FID which were matched with the data of the acoustical description that new person's voice and

the data of the gestalt-description of a face which are given as this result, respectively from these speaker-recognitions section 61 and the face recognition section 62 It is made as [store / in memory 65 / relate with the man's identifier acquired by this dialogue, and].

[0052] moreover, when it judges that the man is a known man, the dialogue control section 63 While making additional study and correction study perform in the speaker-recognition section 61 and the face recognition section 62 by giving an initiation instruction of additional study and correction study to the speaker-recognition section 61 and the face recognition section 62 if needed By carrying out sequential sending out of the predetermined character-string data D2 to predetermined timing with this at the speech synthesis section 64 It is made as [perform / for the speaker-recognition section 61 and the face recognition section 62 to carry out additional study or correction study / dialogue control which is the need and which prolongs a dialogue with the man until the data of an amount are fairly collectable].

[0053] (3) Explain concrete processing of the dialogue control section 63 about an identifier learning function, next the concrete contents of processing of the dialogue control section 63 about an identifier learning function.

[0054] The dialogue control section 63 performs various processings for carrying out sequential study of a new person's identifier according to the identifier study

procedure RT 1 shown in drawing 8 and drawing 9 based on the control program stored in external memory 58 (drawing 5).

[0055] Namely, will start in a step SP 0 and the dialogue control section 63 will set the identifier study procedure RT 1 to the continuing step SP 1, if FID is given from the face recognition section 62 concerned when the face recognition section 62 recognizes people's face based on picture signal S1A from CCD camera 50. The information which associated a known man's identifier stored in memory 65, and SID corresponding to this and FID corresponding to this It judges whether the identifier which is based on (this is associated and it is hereafter called information), and corresponds from the FID can be searched (that is, isn't FID "-1" which means recognition impossible or not?).

[0056] Obtaining an affirmation result in this step SP 1 here means that that man is a known man who the face recognition section 62 has memorized the data of the gestalt-description that man's face, and FID matched with the data concerned relates with that man's identifier, and is stored in memory 65. However, it is also considered also in this case that the face recognition section 62 has incorrect-recognized the new person to be a known man.

[0057] then, when an affirmation result is obtained in a step SP 1, the dialogue control section 63 By progressing to a step SP 2 and sending out the predetermined character-string data D2 to the speech synthesis section 64, as

shown in drawing 10 , he is Mr. "OO. The voice of a question for the identifier of the man " to confirm whether it is in agreement with the identifier (identifier applicable to above-mentioned OO) searched from FID is made to output from a loudspeaker 54.

[0058] Subsequently, the dialogue control section 63 awaits that progress to a step SP 3 and the speech recognition result of the man's "yes and that's right" to this question and the response "no, it is different" is given from the speech recognition section 60. And if this speech recognition result is soon given from the speech recognition section 63 and SID which it is as a result of [at that time] a speaker recognition is given from the speaker-recognition section 61, the dialogue control section 63 will progress to a step SP 4, and will judge whether the man's response is affirmative based on the speech recognition result from the speech recognition section 63.

[0059] It means that it is in the condition that it can be concluded mostly that he is him who has the identifier with which the identifier of obtain [here / in this step SP 4 / an affirmation result] searched based on FID given from the face recognition section 62 in a step SP 1 corresponded with that man's identifier, therefore the dialogue control section 63 searched that man.

[0060] In this way, at this time, the dialogue control section 63 concludes that that man is him who has the identifier which the dialogue control section 63

concerned searched, it progresses to a step SP 5, and an initiation instruction of additional study is given to the speaker-recognition section 61. Moreover, with this, the dialogue control section 63 gives an initiation instruction of correction study, when it gives an initiation instruction of additional study to the speaker-recognition section 61 when SID first given from the speaker-recognition section 61 is in agreement with SID which was stored in memory 65 from this identifier and which is associated and can be searched based on information, and it is not in agreement to this.

[0061] And the dialogue control section 63 progressed to a step SP 6 after this, for example, said like drawing 10 , "It crawls today, and it is and is the weather" etc. If sufficient predetermined time for sequential sending, next addition study, or correction study passes in the speech synthesis section 64, the character-string data D2 for carrying out idle talk for prolonging a dialogue with that man After progressing to a step SP 7 and giving a termination instruction of additional study or correction study to the speaker-recognition section 61 and the face recognition section 62, it progresses to a step SP 20 and the identifier study processing to the man is ended.

[0062] On the other hand, it is a person with the new person by whom face recognition was done in the face recognition section 62 to obtain a negative result in a step SP 1, or it means that the face recognition section 62 has

incorrect-recognized the known man to be a new person. Moreover, it means that the identifier's searched from FID first given from the face recognition section 62 obtaining a negative result in a step SP 4 does not correspond with the man's identifier. And it can be said that it is in the condition that the dialogue control section 63 does not grasp the man correctly in the case of which [these].

[0063] Then, the dialogue control section 63 makes the voice of the question for finding out about the man's identifier "let me know that and an identifier" output from a loudspeaker 54 by progressing to a step SP 8 and giving character-string data D2 to the speech synthesis section 64, as shown in drawing 11 , when a negative result is obtained in a step SP 1, or when a negative result is obtained in a step SP 4.

[0064] And the dialogue control section 63 awaits that progress to a step SP 9 after this, and the speech recognition result (namely, identifier) of response "it is OO" to this question, and the speaker-recognition result (namely, SID) of the speaker-recognition section 61 at the time of the response concerned are given from the speech recognition section 60 and the speaker-recognition section 61, respectively. [of that man]

[0065] And if a speech recognition result is soon given from the speech recognition section 60 and SID is given from the speaker-recognition section 61, the dialogue control section 63 will progress to a step SP 10, and will judge

whether the man is a new person based on FID first given to these speech recognition result and the SID list from the face recognition section 62.

[0066] In the case of the gestalt of this operation, this judgment is made here by the majority of three recognition results which become by the identifier acquired by the speech recognition of the speech recognition section 60, SID from the speaker-recognition section 61, and FID from the face recognition section 62.

[0067] For example, it is judged that it is "-1" as which both FID(s) from SID and the face recognition section 62 from the speaker-recognition section 61 mean recognition impossible, and the man is a new person when the man's identifier which might be based on the speech recognition result from the speech recognition section 60 in step SP is not related with all SID or FID in memory 65. Since it is in the situation that those whom which known face or which known voice does not resemble closely have a completely new identifier, such decision can be performed.

[0068] Moreover, it is judged that the dialogue control section 63 is "-1" as which it is related with the identifier from which FID from SID and the face recognition section 62 from the speaker-recognition section 61 differs in memory 65, or one of these means recognition impossible, and it is a person with the new man also when the man's identifier which might be based on the speech recognition result from the speech recognition section 60 in a step SP 9 is not stored in memory 65.

It is what is easy to happen to incorrect-recognize a new category to be one of the known categories in various recognition processings, and this is because considering that the identifier by which speech recognition was carried out is not registered it can be judged as a new person with quite high reliability.

[0069] On the other hand, the dialogue control section 63 judges that the man is a known man, when the man's identifier which FID from SID and the face recognition section 62 from the speaker-recognition section 61 is related with the same identifier in memory 65, and might be based on the speech recognition result from the speech recognition section 60 in a step SP 9 is an identifier with which the SID and FID were related.

[0070] Moreover, in being the identifier with which either SID which requires the man's identifier which the dialogue control section 63 is related with the identifier from which FID from SID and the face recognition section 62 from the speaker-recognition section 61 differs in memory 65, and might be based on the speech recognition result from the speech recognition section 60 in a step SP 9, or FID was related, the man judges that he is a known man. In this case, since it is thought that the recognition result of either the speaker-recognition section 61 and the face recognition section 62 is wrong, it judges such by this majority.

[0071] On the other hand, the dialogue control section 63 is related with the identifier from which FID from SID and the face recognition section 62 from the

speaker-recognition section 61 differs in memory 65. And when the man's identifier which might be based on the speech recognition result from the speech recognition section 60 in a step SP 9 is an identifier related with neither [this] SID nor FID in memory 65, the man does not judge whether you are a known man or you are a new person. Although it is considered in this case that either the speech recognition section 60, the speaker-recognition section 61 and the face recognition section 62 and all recognition are also wrong, it cannot be judged in this phase. Therefore, this decision is suspended in this case.

[0072] and -- the case where the dialogue control section 63 judges that he is a person with this new man in a step SP 10 by such decision processing -- a step SP 11 -- progressing -- an initiation instruction of new study -- the speaker-recognition section 61 and the face recognition section 62 -- giving -- next -- a step SP 12 -- progressing -- for example, drawing 11 -- like -- "-- I am a robot. thank you for your consideration. " -- or -- "-- Mr. OO and today -- it's a nice day, isn't it. " -- etc. -- the character-string data D2 for carrying out idle talk which prolongs a dialogue with the man are sent out to the speech synthesis section 64.

[0073] Moreover, it repeats step SP12-SP13-SP's12 loop formation until it will return to a step SP 12 and will obtain an affirmation result in a step SP 13 after this, if the dialogue control section 63 judges whether it progressed to a step SP 13 after this, and both collection of the data of the acoustical description in the

speaker-recognition section 61 and collection of the data of the gestalt-description of the face in the face recognition section 62 reached the amount enough and a negative result is obtained.

[0074] And if an affirmation result is obtained in a step SP 13 when both collection of the data of the acoustical description in the speaker-recognition section 61 and collection of the data of the gestalt-description of the face in the face recognition section 62 reach an amount enough soon, the dialogue control section 63 will progress to a step SP 14, and will give a termination instruction of new study to these speaker-recognitions section 61 and the face recognition section 62. Consequently, in the speaker-recognition section 61, the data of that acoustical description are matched with new SID, and are memorized, and in the face recognition section 62, the data of that gestalt-description are matched with new FID, and are memorized.

[0075] Moreover, it relates with the identifier of that man that might be based [in / when it awaits that the dialogue control section 63 progresses to a step SP 15 after this, and this SID and FID are given from the speaker-recognition section 61 and the face recognition section 62 respectively and these are given soon, as it is shown, for example in drawing 12 / a step SP 9] on the speech recognition result from the speech recognition section 60 in these, and registers with memory 65. And the dialogue control section 63 progresses to a step SP 20 after

this, and ends the identifier study processing to that man.

[0076] on the other hand, when it is judged in a step SP 10 that this man is a known man, the dialogue control section 63 When it progressed to a step SP 16 and the speaker-recognition section 61 and the face recognition section 62 can recognize the known man correctly (that is, the speaker-recognition section 61 and the face recognition section 62) When same SID or same SID as SID corresponding to the known man stored in memory 65 as correlation information or FID is being outputted as a recognition result An initiation instruction of additional study is given to the speaker-recognition section 61 or the face recognition section 62. When the speaker-recognition section 61 and the face recognition section 62 have not recognized the known man correctly (that is, the speaker-recognition section 61 and the face recognition section 62) When same SID or same SID as SID corresponding to the known man stored in memory 65 as correlation information or FID is being outputted as a recognition result, an initiation instruction of correction study is given to the speaker-recognition section 61 or the face recognition section 62.

[0077] SID from the speaker-recognition section 61 from which the dialogue control section 63 was specifically obtained in a step SP 9, FID first given from the face recognition section 62 is related with the same identifier in memory 65. and when the identifier which might be based on the speech recognition result

from the speech recognition section 60 in a step SP 9 is an identifier with which the SID and FID were related and the man judges that he is a known man in a step SP 10 An initiation instruction of additional study is given to the speaker-recognition section 61 and the face recognition section 62, respectively.

[0078] Moreover, SID from the speaker-recognition section 61 from which the speaker-recognition section 63 was obtained in a step SP 9, It is related with the identifier from which FID first given from the face recognition section 62 differs in memory 65. and when the man judges that he is a known man in a step SP 10 by being the identifier with which either SID which requires the identifier which might be based on the speech recognition result from the speech recognition section 60 in a step SP 9, or FID was related While SID related with the identifier which might be based on the speech recognition result from the speech recognition section 60, or FID was outputted gives an initiation instruction of additional study to the speaker-recognition section 61 or the face recognition section 62. An initiation instruction of correction study is given to the face recognition section 62 or the speaker-recognition section 61 of another side which outputted FID which is not related with the identifier which might be based on the speech recognition result from the speech recognition section 60, or SID.

[0079] And the dialogue control section 63 should move on to a step SP 17 after this, for example, should be shown in drawing 13 . "oh, he is Mr. OO.

Recollections stripes were carried out. today -- it's a nice day, isn't it. " and "last time -- growing - when -- meeting now -- the bottom -- ****. Sequential sending " -- etc. -- the character-string data D2 for carrying out idle talk for prolonging a dialogue with the man in the speech synthesis section 64 If predetermined time sufficient after this for additional study or correction study passes, after progressing to a step SP 18 and giving a termination instruction of additional study or correction study to the speaker-recognition section 61 and the face recognition section 62, it progresses to a step SP 20 and the identifier study processing to that man is ended.

[0080] on the other hand, when it is judged that it cannot judge with the dialogue control section 63 being a new person in a step SP 10 as this man is a known man, it progresses to a step SP 19, for example, is shown in drawing 14 -- as -- "-- such -- really. Are you fine? Sequential sending out of the character-string data D2 for carrying out idle talk of " etc. is carried out at the speech synthesis section 64.

[0081] And in this case, it will progress to a step SP 20 and the dialogue control section 63 will end the identifier study processing to that man, if an initiation instruction and its termination instruction of new study and addition study or correction study are not given to the speaker-recognition section 61 and the face recognition section 62 (that is, both new study and addition study and correction

study are not made to perform in the speaker-recognition section 61 and the face recognition section 62) but predetermined time passes.

[0082] Thus, the dialogue control section 63 is made as [carry out / sequential study of a new person's identifier] based on each recognition result of the speech recognition section 60, the speaker-recognition section 61, and the face recognition section 62 by performing dialogue control with people, and motion control of the speaker-recognition section 61 and the face recognition section 62.

[0083] (4) Explain the concrete configuration of the speech recognition section 60 for embodying the concrete configuration, next the above identifier learning functions of the speech recognition section 60 and the face recognition section 62, and the face recognition section 62.

[0084] (4-1) The concrete block diagram 15 of the speech recognition section 60 shows the concrete configuration of this speech recognition section 60.

[0085] In this speech recognition section 60, sound signal S1B from a microphone 51 is inputted into the AD (Analog Digital) transducer 70. The AD translation section 70 samples sound signal S1B which is the analog signal supplied, quantizes, and carries out A/D conversion to the voice data which is a digital signal. This voice data is supplied to the feature-extraction section 71.

[0086] The feature-extraction section 71 performs for example, MFCC (Mel Frequency Cepstrum Cofficient) analysis for every suitable frame about the

voice data inputted there, and outputs MFCC obtained as a result of the analysis to the matching section 72 and the non-registered word section processing section 76 as a feature vector (feature parameter). In addition, in the feature-extraction section 71, it is possible to extract after that, for example, linear predictor coefficients, a cepstrum multiplier, a line spectrum pair, the power (output of a filter bank) for every predetermined frequency, etc. as a feature vector.

[0087] the voice (input voice) inputted into the microphone 51 while the matching section 72 referred to the sound model storage section 73, the dictionary storage section 74, and the syntax storage section 75 if needed using the feature vector from the feature-extraction section 71 -- for example, the continuous distribution HMM (Hidden Markov Model) -- speech recognition is carried out based on law.

[0088] That is, the sound model storage section 73 has memorized the sound model (for example, the standard pattern used for DP (Dynamic Programing) matching besides HMM is included) showing the description acoustical about subWORD, such as each phoneme in the audio language which carries out speech recognition, and syllable, a phoneme. in addition -- here -- continuous distribution HMM -- since it is carrying out performing speech recognition based on law, HMM (Hidden Markov Model) is used as a sound model.

[0089] The dictionary storage section 74 recognizes the word dictionary in which

the information (sound information) about the pronunciation of the word clustered for every unit for recognition and the header of the word were matched.

[0090] Here, drawing 16 shows the word dictionary memorized by the dictionary storage section 74.

[0091] As shown in drawing 16 , in the word dictionary, the header and its phoneme sequence of a word are matched and the phoneme sequence is clustered for every corresponding word. In the word dictionary of drawing 16 , one entry (one line of drawing 16) is equivalent to one cluster.

[0092] In addition, in drawing 16 , the header is expressed in a Roman alphabet and Japanese (kana kanji), and the phoneme sequence is expressed with the Roman alphabet. However, "N" in a phoneme sequence expresses a syllabic nasal "**." Moreover, in drawing 16 , although one phoneme sequence is described to one entry, it is also possible to describe two or more phoneme sequences to one entry.

[0093] Return and the syntax storage section 26 have memorized the syntax rule which described how it was that each word registered into the word dictionary of the dictionary storage section 25 carries out a chain (connected) to drawing 4 .

[0094] Here, drawing 17 shows the syntax rule memorized by the syntax storage section 75. In addition, the syntax rule of drawing 17 is described by EBNF (Extended Backus Naur Form).

[0095] In drawing 17 , even";" which appears in the beginning from the head of the sentence expresses one syntax rule. Moreover, the alphabet (train) by which "$" was given to the head expresses a variable, and the alphabet (train) to which "$" is not given expresses the header (header in the Roman alphabet shown in drawing 16) of a word. Meaning that the part furthermore surrounded by [] can be omitted, "|" means choosing either of the words (or variable) of the header arranged before and behind that.

[0096] therefore, drawing 17 -- setting -- for example, the syntax rule "$col=[Kono|sono] iro wa;" of the 1st line (from a top to the 1st line) -- variable $col -- "-- this -- be (color) -- " -- or -- "-- the -- be (color) -- " -- ** -- it expresses that it is the word train to say.

[0097] In addition, in the syntax rule shown in drawing 17 , although variable $sil and $garbage are not defined, variable $sil expresses a silent sound model (silent model), and, fundamentally, variable $garbage expresses the GABEJI model which permitted the free transition between phonemes.

[0098] Again, in drawing 15 , by referring to the word dictionary of the dictionary storage section 74, return and the matching section 72 are connecting the sound model memorized in the sound model storage section 73, and constitute the sound model (word model) of a word. the word model which connected the matching section 72 by referring to the syntax rule memorized by the syntax

storage section 75 in some word models, and was furthermore connected by making it such -- using -- a feature vector -- being based -- continuous distribution HMM -- the voice inputted into the microphone 51 is recognized by law. That is, the matching section 72 detects the sequence of a word model with the most expensive score (likelihood) with which the feature vector of the time series which the feature-extraction section 71 outputs is observed, and outputs the header of the word train corresponding to the sequence of the word model as a recognition result of audio.

[0099] the word model which connected the matching section 72 with the word corresponding to the connected word model, and was more specifically connected by making it such -- using -- a feature vector -- being based -- continuous distribution HMM -- the voice inputted into the microphone 51 is recognized by law. That is, the matching section 72 detects the sequence of a word model with the most expensive score (likelihood) with which the feature vector of the time series which the feature-extraction section 71 outputs is observed, and outputs the header of the word train corresponding to the sequence of the word model as a speech recognition result.

[0100] More specifically, the matching section 72 outputs the header of the word train which accumulates the appearance probability (output probability) of each feature vector, and makes the score the highest by using the accumulation value

as a score about the word train corresponding to the connected word model as a speech recognition result.

[0101] The speech recognition result inputted into the microphone 51 outputted as mentioned above is outputted to the dialogue control section 63 as character-string data D1.

[0102] Although there is syntax rule (suitably henceforth the regulation for non-registered words) "$pat1=$colorl $garbage $color2;" using variable $garbage which expresses a GABEJI model to the 9th line (from a top to the 9th line) with the gestalt of operation of drawing 17 here The matching section 72 detects the voice section corresponding to variable $garbage as the voice section of a non-registered word, when [this] it sees and the regulation for registration words is applied. Furthermore, the matching section 72 detects the phoneme sequence as transition of the phoneme in the GABEJI model which variable $garbage when the regulation for non-registered words is applied expresses as a phoneme sequence of a non-registered word. And the matching section 72 supplies the voice section and the phoneme sequence of a non-registered word which are detected when the speech recognition result to which the regulation for non-registered words was applied is obtained to the non-registered word section processing section 76.

[0103] In addition, according to the above-mentioned regulation for

non-registered words "$pat1=$colorl $garbage $color2;" Although one non-registered word between the phoneme sequence of the word (train) registered into the word dictionary expressed with variable #color1 and the phoneme sequence of the word (train) registered into the word dictionary expressed with variable $color2 is detected In the gestalt of this operation, even if it is the case where two or more non-registered words are included in utterance, and the case where it is not inserted between the words (train) by which the non-registered word is registered into the word dictionary, it is applicable.

[0104] The non-registered word section processing section 76 stores temporarily the sequence (feature-vector sequence) of the feature vector supplied from the feature-extraction section 71. Furthermore, the non-registered word section processing section 76 will detect the feature-vector sequence of the voice in the voice section from the feature-vector sequence stored temporarily, if the voice section and the phoneme sequence of a non-registered word are received from the matching section 72. And the non-registered word section processing section 76 gives unique ID (identification) to a phoneme sequence (non-registered word) from the matching section 72, and supplies it to the feature-vector buffer 77 with the phoneme sequence of a non-registered word, and the feature-vector sequence in the voice section.

[0105] The feature-vector buffer 77 matches and stores temporarily ID, phoneme

sequence, and feature-vector sequence of the non-registered word supplied from the non-registered word section processing section 76, as shown in drawing 18.

[0106] In drawing 18 , several sequential 00 from 1 are attached as ID to the non-registered word here. [when it follows, for example, ID, phoneme sequence, and feature-vector sequence of a non-registered word of N individual are now memorized in the feature-vector buffer 77] When the matching section 72 detects the voice section and the phoneme sequence of a non-registered word, in the non-registered word section processing section 76 N+1 is attached as ID to the non-registered word, and in the feature-vector buffer 77, as a dotted line shows to drawing 18 , ID, phoneme sequence, and feature-vector sequence of the non-registered word are memorized.

[0107] Return and the clustering section 78 calculate the score to each of other already memorized non-registered word (suitably henceforth the word non-registered [memorized]) to the feature-vector buffer 77 again at drawing 15 about the non-registered word (suitably henceforth a new sheep registration word) newly memorized by the feature-vector buffer 77.

[0108] That is, the clustering section 78 carries out [voice / input] a new sheep registration word, and considers that the word non-registered [memorized] is the word registered into the word dictionary, and calculates the score to each

****** non-registered language about a new sheep registration word as well as the case in the matching section 72. The clustering section 78 connects a sound model according to the phoneme sequence of the word non-registered [memorized], and, specifically, calculates the score as likelihood with which the feature-vector sequence of a new sheep registration word is observed from the connected sound model while it recognizes the feature-vector sequence of a new sheep registration word by referring to the feature-vector buffer 77.

[0109] In addition, that the sound model is remembered to be by the sound model storage section 73 is used.

[0110] Similarly, the clustering section 78 also calculates the score to a new sheep registration word about each ****** non-registered language, and updates the score sheet memorized by the score sheet storage section 79 with the score.

[0111] Furthermore, the clustering section 78 detects the cluster which adds a new sheep registration word as a new member by referring to the updated score sheet out of the cluster which clustered the non-registered word (word non-registered [memorized]) already searched for. Furthermore, the clustering section 78 considers as the new member of the cluster which detected the new sheep registration word, divides the cluster based on the member of the cluster, and updates the score sheet memorized by the score sheet storage section 79 based on the division result.

[0112] The score sheet storage section 79 memorizes the score sheet with which the score to the word about a new sheep registration word non-registered [memorized], the score to the new sheep registration word about the word non-registered [memorized], etc. were registered.

[0113] Here, drawing 19 shows the score sheet.

[0114] A score sheet consists of entries "ID", a "phoneme sequence", a "cluster number", the "representation member ID", and a "score" of a non-registered word were described to be.

[0115] The thing same as "ID" and a "phoneme sequence" of a non-registered word as what was memorized by the feature-vector buffer 77 is registered by the clustering section 78. A "cluster number" is a figure for specifying the cluster from which the non-registered word of the entry serves as a member, is attached by the clustering section 78 and registered into a score sheet. "The representation number ID" is non-registered ID as a representation member representing the cluster from which the non-registered word of that entry serves as a member, and can recognize the representation member of the cluster from which the non-registered word serves as a member by this representation member ID. In addition, the representation member of a cluster is called for by the clustering section 29, and ID of the representation member is registered into the representation member ID of a score sheet. A "score" is a score to each of

other non-registered word about the non-registered word of the entry, and as mentioned above, it is calculated by the clustering section 78.

[0116] For example, now, in the feature-vector buffer 77, supposing ID, phoneme sequence, and feature-vector sequence of a non-registered word of N individual are memorized, ID of the non-registered word of the N individual, the phoneme sequence, the cluster number, the representation number ID, and the score are registered into the score sheet.

[0117] And in the clustering section 78, if ID, phoneme sequence, and feature-vector sequence of a new sheep registration word are newly memorized by the feature-vector buffer 77, as a score sheet shows drawing 19 by the dotted line, it will be updated by it.

[0118] Namely, a score [as opposed to each word about ID of a new sheep registration word, a phoneme sequence, a cluster number, the representation member ID, and a new sheep registration word non-registered / memorized / in a score sheet] (the scores s (1 N+ 1) and s (2 N+1) in drawing 19 and --s (N+1, N) are added.) Furthermore, the score (s (1 N+ 1) in drawing 19 , s (2 N+1), --s (N+1, N)) to the new sheep registration word about each word non-registered [memorized] is added to a score sheet. The cluster number of a non-registered word and the representation member ID in a score sheet are changed if needed so that it may furthermore mention later.

[0119] In addition, in the gestalt of operation of drawing 19 , ID has expressed the score [as opposed to the non-registered word (phoneme sequence) of j in ID] about the non-registered word (utterance) of i as s (i, j).

[0120] Moreover, the score [as opposed to / ID / the non-registered word (phoneme sequence) of i in ID about the non-registered word (utterance) of i] s (i, j) is registered into a score sheet (drawing 19). However, in the matching section 72, since this score s (i, j) is calculated when the phoneme sequence of a non-registered word is detected, it is not necessary to calculate it in the clustering section 78.

[0121] Return and the maintenance section 80 update again the word dictionary memorized by the score sheet in the dictionary storage section 74 based on the score sheet after updating in the storage section 79 to drawing 15 .

[0122] Here, the representation member of a cluster is determined as follows. That is, let what makes max total (the average value which did the division of the total by the number of other non-registered words is sufficient in addition to this) of the score about each of other non-registered word, for example among the non-registered words used as the member of a cluster be the representation member of the cluster. Therefore, when it is expressing with k the member ID of the member which belongs to a cluster in this case, it is a degree type [0123].

[Equation 1]

$$K=\max_{k}\left\{\sum s(k^{3},k)\right\}$$

[0124] The member which comes out and sets the value k (**k) shown to ID will be made into a representation member.

[0125] however, (1) type -- setting -- maxk{} -- {} -- k which makes an inner value max is meant. Moreover, k3 means ID of the member belonging to a cluster like k. Furthermore, sigma means the total to which k3 continues and changes [make] to ID of all the members belonging to a cluster.

[0126] In addition, when determining a representation member as mentioned above and the member of a cluster is 1 or two non-registered words, in deciding a representation member, it is not necessary to calculate a score. That is, when the member of a cluster is one non-registered word, the one non-registered word serves as a representation member, and when the members of a cluster are two non-registered words, it is good also considering any of the two non-registered words as a representation member.

[0127] Moreover, the decision approach of a representation member can also make into the representation member of the cluster what makes min total of the distance in feature-vector space with each of other non-registered word among

the non-registered words which are not limited to what was mentioned above and serve as a member of a cluster.

[0128] In the speech recognition section 60 constituted as mentioned above, speech recognition processing which recognizes the voice inputted into the microphone 51, and non-registered word processing about a non-registered word are performed according to the speech recognition procedure RT 2 shown in drawing 20 .

[0129] In practice, in the speech recognition section 60, if the feature-extraction section 71 is given it, sound signal S1B obtained when people spoke being used as voice data through the AD translation section 70 from a microphone 51, this speech recognition procedure RT 2 will be started in a step SP 30.

[0130] And in the continuing step SP 31, by carrying out sonagraphy of the voice data per predetermined frame, the feature-extraction section 71 extracts a feature vector, and supplies the sequence of the feature vector to the matching section 72 and the non-registered word section processing section 76.

[0131] In continuing step S32, about the special order OBEKUTORU sequence from the feature-extraction section 71, the matching section 76 performs score count, as mentioned above, and it outputs it in quest of the header of the word train which brings a speech recognition result in step S33 based on the score obtained as a result of score count after this.

[0132] Furthermore, the matching section 72 judges whether the non-registered word was included in a user's voice in continuing step S34.

[0133] Here, in this step S34, when judged with the non-registered word not being included in a user's voice (i.e., when a speech recognition result is obtained without applying the above-mentioned regulation for non-registered words "$pat1=$colorl $garbage $color2;"), it progresses to step S35 and processing is completed.

[0134] On the other hand, when judged with the non-registered word being included in a user's voice in step S34, When the regulation for non-registered words "$pat1=$colorl $garbage $color2;" is applied and a speech recognition result is obtained, namely, the matching section 23 In continuing step S35, while detecting the voice section corresponding to variable $garbage of the regulation for non-registered words as the voice section of a non-registered word The phoneme sequence as transition of the phoneme in the GABEJI model which the variable $garbage expresses is detected as a phoneme sequence of a non-registered word, the voice section and the phoneme sequence of the non-registered word are supplied to the non-registered word section processing section 76, and processing is ended (step SP 36).

[0135] On the other hand, the non-registered word engine processing section 76 will detect the feature-vector sequence of the voice in the voice section, if the

feature-vector sequence supplied from the feature-extraction section 71 is stored temporarily and the voice section and the phoneme sequence of a non-registered word are supplied from the matching section 72. Furthermore, the non-registered word section processing section 76 gives ID to a non-registered word (phoneme sequence) from the matching section 72, and supplies it to the feature-vector buffer 77 with the phoneme sequence of a non-registered word, and the feature-vector sequence in the voice section.

[0136] If ID, phoneme sequence, and feature-vector sequence of a new non-registered word (new sheep registration word) are memorized by the feature-vector buffer 77 as mentioned above, processing of a non-registered word will be performed after this according to the non-registered word-processing procedure RT 3 shown in drawing 21 .

[0137] That is, in the speech recognition section 60, if ID, phoneme sequence, and feature-vector sequence of a new non-registered word (new sheep registration word) are memorized by the feature-vector buffer 77 as mentioned above, this non-registered word-processing procedure RT 3 will be started in a step SP 40, and in step S41, the clustering section 78 reads ID and the phoneme sequence of a new sheep registration word from the feature-vector buffer 77 first.

[0138] Subsequently, in step S42, when the clustering section 78 refers to the

score sheet of the score sheet storage section 30, it judges whether the cluster already called for (generated) exists.

[0139] and the cluster already called for in this step S42, when judged with not existing namely, when a new sheep registration word is first non-registered word and the entry of the word non-registered [memorized] does not exist in a score sheet The information progress to step S43, and the clustering section 78 newly generates the cluster which makes the new sheep registration word a representation member, and concerning the new cluster, A score sheet is updated by registering the information about a kind registration word into the score sheet of the score sheet storage section 79.

[0140] That is, the clustering section 78 registers into a score sheet (drawing 19) ID and the phoneme sequence of a new sheep registration word which were read from the feature-vector buffer 77. Furthermore, the clustering section 78 generates a unique cluster number, and registers it into a score sheet as a cluster number of a new sheep registration word. Moreover, the clustering section 78 registers ID of a new sheep registration word into a score sheet as a representation number ID of the new sheep registration word. Therefore, a new sheep registration word serves as a representation member of a new cluster in this case.

[0141] In addition, since the word which calculates a score with a new sheep

registration word non-registered [memorized] does not exist in now, count of a score is not performed.

[0142] After processing of this step S43 progresses to step S52, and based on the score sheet updated at step S43, the maintenance section 80 updates the word dictionary of the dictionary storage section 74, and ends processing (step SP 54).

[0143] That is, since a new cluster is generated in now, the maintenance section 31 recognizes the newly generated cluster with reference to the cluster number in a score sheet. And the maintenance section 80 adds the entry corresponding to the cluster to the word dictionary of the dictionary storage section 74, and registers the phoneme sequence of a new sheep registration word the new phoneme sequence of the representation member of a cluster, i.e., the case of now, as a phoneme sequence of the entry.

[0144] When judged with on the other hand the cluster already called for existing in step S42, A new sheep registration word is not first non-registered word. Namely, on therefore, a score sheet (drawing 19) When the entry (line) of the word non-registered [memorized] exists, it progresses to step S44, and the clustering section 78 calculates the score to a new sheep registration word about each ****** each of non-registered language while calculating the score to each ****** each of non-registered language about a new sheep registration word.

[0145] When the word of 1 thru/or N individual non-registered [memorized] exists and ID, for example, sets ID of a new sheep registration word to N+1 now, namely, in the clustering section 78 The scores s (1 N+ 1), s (1 N+ 2), --, s (N, N+1) to each word of N individual about the new sheep registration word of the part shown by the dotted line in drawing 19 non-registered [memorized], The scores s (1 N+1), s (2 N+1), --, s (N, N+1) to the new sheep registration word about each word of N individual non-registered [memorized] are calculated. In addition, in the clustering section 78, in calculating these scores, the feature-vector sequence of a new sheep registration word and each word of N individual non-registered [memorized] is needed, but these feature-vector sequences are recognized by referring to the feature-vector buffer 28.

[0146] And the clustering section 78 adds the calculated score to a score sheet (drawing 19) with ID and the phoneme sequence of a new sheep registration word, and progresses to step S45.

[0147] At step S45, the clustering section 78 detects the cluster which has the representation member which makes the highest (greatly) the score s about a new sheep registration word (N+1, i) (i= 1, 2, --, N) by referring to a score sheet (drawing 19). That is, by referring to the representation member ID of a score sheet, the clustering section 78 recognizes the word used as a representation member non-registered [memorized], is referring to the score of a score sheet

further, and detects the word as a representation member which makes the highest the score about a new sheep registration word non-registered [memorized]. And the clustering section 78 detects the cluster of the cluster number of the word as the detected representation member non-registered [memorized].

[0148] Then, it progresses to step S46 and the clustering section 29 is added to the member of the cluster (suitably henceforth a detection cluster) which detected the new sheep registration word at step S45. That is, the clustering section 78 writes in the cluster number of the representation member of a detection cluster as a cluster number of the new sheep registration word in a score sheet.

[0149] And in step S47, the clustering section 78 performs cluster division processing in which a detection cluster is divided into two clusters, and progresses to step S48. At step S48, by cluster division processing of step S47, the clustering section 78 judges whether the detection cluster was able to be divided into two clusters, and when [which was able to be divided] it judges, it progresses to step S49. At step S49, the clustering section 78 finds the distance between clusters between two clusters (these two clusters are hereafter called suitably the 1st child cluster and 2nd child cluster) obtained by division of a detection cluster.

[0150] Here, it is defined as the distance between clusters between the 1st and 2nd child clusters as follows, for example.

[0151] That is, while expressing ID of the member (non-registered word) of the arbitration of both the 1st child cluster and the 2nd child cluster with k, when ID of the representation member (non-registered word) of the 1st and 2nd child cluster is expressing with k1 or k2, respectively, it is a degree type [0152].

[Equation 2]

$$D(k1, k2)=\mathrm{maxval}_{k}\left\{\mathrm{abs}(\log(s(k, k1))-\log(\right.$$

[0153] It comes out and let the value D (k1, k2) expressed be the distance between clusters between the 1st and 2nd child cluster.

[0154] However, in (2) types, abs() expresses the absolute value of the value in (). moreover, maxvalk -- {-- {} which} changes k and is called for -- the maximum of an inner value is expressed. Moreover, log expresses a natural logarithm or a common logarithm.

[0155] When it is that ID expresses the member of i as member #I now, it is equivalent to the distance of member #k and the representation member k1, and 1/s (k, k1) of inverse numbers of the score in (2) types is equivalent to the

distance of member #k and the representation member k2 1/s (k, k2) of inverse numbers of a score. Therefore, according to the (2) types, maximum of the distance of representation member #k1 of the 1st child cluster and the difference of representation member #k2 of the 2nd child cluster will be made into the child distance between clusters between the 1st and 2nd child cluster among the members of the 1st and 2nd child cluster.

[0156] In addition, by not being limited to what was mentioned above and performing DP matching of the representation member of the 1st child cluster, and the representation member of the 2nd child cluster in addition, the addition value of the distance in feature-vector space is calculated, and distance between clusters can also make the addition value of the distance into distance between clusters.

[0157] After processing of step S49, it progresses to step S50 and, as for the clustering section 78, the distance between cluster ** of the 1st and 2nd child cluster judges whether it is size (or it is beyond the threshold xi) from the predetermined threshold xi.

[0158] In step S50, when it judges that distance between clusters is size from the predetermined threshold xi (i.e., when it is thought that two or more un-registering-as member of detection cluster back is what should say from the acoustical description and should be clustered to two clusters), it progresses to

step S51 and the clustering section 78 registers the 1st and 2nd child cluster into the score sheet of the score sheet storage section 79.

[0159] That is, the clustering section 78 assigns a unique cluster number to the 1st and 2nd child cluster, and although clustered by the 2nd child cluster, it updates a score sheet so that a cluster number may be made into the cluster number of the 2nd child cluster, while it makes a cluster number the cluster number of the 1st child cluster, although it was clustered by the 1st child cluster among the members of a detection cluster.

[0160] Furthermore, the clustering section 78 updates a score sheet so that the representation member ID of the member clustered by the 2nd child cluster may be set to ID of the representation member of the 2nd child cluster, while setting to ID of the representation member of the 1st child cluster the representation member ID of the member clustered by the 1st child cluster.

[0161] In addition, it is possible to assign either KU r SUTANAMBA of a detection cluster among the 1st and 2nd child cluster.

[0162] If the clustering section 78 registers the 1st and 2nd child cluster into a score sheet as mentioned above, it progresses to S52 from step S51, and based on a score sheet, the maintenance section 80 will update the word dictionary of the dictionary storage section 74, and will end processing (step SP 54).

[0163] That is, since the detection cluster was divided into the 1st and 2nd child

cluster in now, the maintenance section 80 deletes the entry corresponding to the detection cluster in a word dictionary first. Furthermore, the maintenance section 80 adds two entries corresponding to each 1st and 2nd child cluster to a word dictionary, and as a phoneme sequence of the entry corresponding to the 1st child cluster, it registers the phoneme sequence of the representation member of the 2nd child cluster as a phoneme sequence of the entry corresponding to the 2nd child cluster while it registers the phoneme sequence of the representation member of the 1st child cluster.

[0164] On the other hand, it sets to step S48. By cluster division processing of step S47 In the step S50 when judged with the ability of a detection cluster to have not been divided into two clusters When it judges that the distance between clusters of the 1st and 2nd child cluster is not size from the predetermined threshold xi, Therefore, when it is not what is not alike, so that the acoustical description after two or more un-registering as a member of a detection cluster clusters to the 1st and 2nd child cluster, Progressing to step S53, the clustering section 78 asks for the new representation member of a detection cluster, and updates a score sheet.

[0165] That is, the clustering section 78 recognizes the score s required for count of (1) type (k3, k) by referring to the score sheet of the score sheet storage section 79 about each member of the detection cluster which added the new

sheep registration back as a member. Furthermore, clustering 78 calculates ID of the new representation member of a detection cluster, and the becoming member based on (1) type using the recognized score s (k3, k). And the clustering section 78 rewrites the representation member ID of each member of the detection cluster in a score sheet (drawing 19) to ID of the new representation member of a detection cluster.

[0166] Then, it progresses to step S52, and the maintenance section 80 updates the word dictionary of the dictionary storage section 74 based on a score sheet, and ends processing (step SP 54).

[0167] That is, in now, by referring to a score sheet, the maintenance section 80 recognizes the new representation member of a detection cluster, and recognizes the phoneme sequence of the DA table member further. And the maintenance section 80 changes the phoneme sequence of the entry corresponding to the detection cluster in a word dictionary into the phoneme sequence of the new representation member of a detection cluster.

[0168] Here, cluster division processing of the step SP 47 of drawing 21 is performed according to the cluster division procedure RT 4 shown in drawing 22 .

[0169] That is, in the speech recognition processing section 60, if it progresses to a step SP 47 from the step SP 46 of drawing 22 , this cluster division procedure RT 4 is started in a step SP 60, and first, in step S61, the clustering

section 78 will choose the combination of two members of the arbitration which has not been chosen yet, and will make each a temporary representation member from the detection cluster to which the new sheep registration back was added as a member. Here, these two temporary representation members are hereafter called suitably the 1st temporary representation member and 2nd temporary representation member.

[0170] And in continuing step S62, the clustering section 78 judges whether the member of a detection cluster can be divided into two clusters so that the 1st temporary representation member and the 2nd temporary representation member can be made into a representation member, respectively.

[0171] Here, although whether the 1st or 2nd temporary representation member can be made into a representation member needs to calculate (1) type, the score s (k', k) used for this count is recognized by referring to a score sheet.

[0172] In step S62, when it is judged with the ability of the member of a detection cluster not to be divided into two clusters so that the 1st temporary representation member and the 2nd temporary representation member can be made into a representation member, respectively, step S62 is skipped and it progresses to step S64.

[0173] Moreover, in step S62, so that the 1st temporary representation member and the 2nd temporary representation member can be made into a

representation member, respectively When judged with the ability of the member of a detection cluster to be divided into two clusters, it progresses to step S63. The clustering section 78 So that the 1st temporary representation member and the 2nd temporary representation member may turn into a representation member, respectively The member of a detection cluster is divided into two clusters, and it progresses to step S64 as a candidate (suitably henceforth the group of a candidate cluster) of the 1st [which brings a division result of a detection cluster in the group of two clusters after the division], and 2nd child clusters.

[0174] At step S64, the clustering section 78 It judges whether in the member of a detection cluster, there is any group of two members which have not been chosen as a group of the 1st and 2nd temporary representation member yet. When it judges with it being, the group of two members, return and the detection cluster which has not been chosen as a group of the 1st and 2nd temporary representation member yet, is chosen as step S61, and the same processing is repeated hereafter.

[0175] Moreover, in step S64, when judged with there being no group of two members of the detection cluster which has not been chosen as a group of the 1st and 2nd temporary representation member, it progresses to step S65 and the clustering section 78 judges whether the group of a candidate cluster exists.

[0176] In step S65, when judged with the group of a candidate cluster not existing, the return of step S66 is skipped and carried out. In this case, in step S48 of drawing 21 , it is judged with the ability of a detection cluster to have not been divided.

[0177] On the other hand, when judged with the group of a candidate cluster existing in step S65, it progresses to step S66, and the clustering section 78 finds the distance between clusters between two clusters of the group of each candidate cluster, when two or more groups of a candidate cluster exist. And distance between clusters asks for the group of the minimum candidate cluster, and carries out the division result of a detection cluster for the group of the candidate cluster, namely, the clustering section 78 carries out a return to the 1st as 2nd child cluster. In addition, only as for one case, let the group of the candidate cluster be the 1st and 2nd child cluster for the group of a candidate cluster as it is.

[0178] In this case, in step S48 of drawing 21 , it is judged with the ability of the detection cluster to have been divided.

[0179] as mentioned above, out of the cluster which clustered the non-registered word already searched for in the clustering section 78 Detect the cluster (detection cluster) which adds a new sheep registration word as a new member, and since the detection cluster was divided as a new member of the detection

cluster based on the member of the detection cluster, a new sheep registration word It can cluster easily to those to which the acoustical description approximates the non-registered word.

[0180] Registration to the word dictionary of a non-registered word can be performed easily, avoiding large-scale-ization of a word dictionary in the maintenance section 80, furthermore, since the word dictionary was updated based on such a clustering result.

[0181] Moreover, temporarily, even if it mistakes detection of the voice section of a non-registered word in the matching section 72, for example, such a non-registered word is clustered by division of a detection cluster at a cluster different from the non-registered word by which the voice section was detected correctly. And although the entry corresponding to such a cluster will be registered into a word dictionary, since the phoneme sequence of this entry becomes a thing corresponding to the voice section which was not detected correctly, it does not give a big score in subsequent speech recognition. Therefore, even if it mistakes detection of the voice section of a non-registered word, the error will hardly influence subsequent speech recognition.

[0182] Here, drawing 23 shows the clustering result obtained by uttering a non-registered word. In addition, in drawing 23 , each entry (each line) expresses one cluster. Moreover, the left column of drawing 23 expresses the

phoneme sequence of the representation member (non-registered word) of each cluster, and the right column of drawing 23 expresses the contents of utterance and the number of non-registered words used as the member of each cluster.

[0183] That is, in drawing 23 , the entry of the 1st line expresses the cluster from which only one utterance of a non-registered word "a bath" serves as a member, and the phoneme sequence of the representation member has become "doroa:" (drawer). Moreover, the entry of the 2nd line expresses the cluster from which three utterance of a non-registered word "a bath" serves as a member, for example, and the phoneme sequence of the representation member has become "kuro" (clo).

[0184] Furthermore, the entry of the 7th line expresses the cluster from which four utterance of a non-registered word "a book" serves as a member, for example, and the phoneme sequence of the representation member has become "NhoNde:su" (NHONTESU). Moreover, the entry of the 8th line expresses the cluster from which one utterance of a non-registered word "Orange" and utterance of a non-registered word "a book" of 19 serve as a member, for example, and the phoneme sequence of the representation member has become "ohoN" (OHON). Other entries express the same thing.

[0185] According to drawing 23 , it turns out about utterance of the same non-registered word that it is clustered good.

"GABOA filtering (Gabor Filtering)" as face extract processing in which a face pattern is extracted, and a face is recognized from a face pattern.

[0189] This face recognition section 62 has the study phase where the face recognition processing section 91 learns a face pattern, and the recognition phase of recognizing the face pattern extracted from picture signal S1A based on the learned data.

[0190] The structure of the study phase of the face recognition section 62 is shown in drawing 24 , and the configuration of the recognition phase of the face recognition section 62 is shown in drawing 25 , respectively.

[0191] In a study phase, as shown in drawing 24 , it is supplied to the face recognition processing section 91 which the result of having carried out the face extract of a user's image pick-up image inputted from CCD camera 50 (drawing 5) in the face extract processing section 90 which becomes with a GABOA filter becomes with a support vector machine. In the face recognition processing section 91, a provisional discriminant function is obtained using the data, i.e., the teacher data, for study supplied from the outside.

[0192] Moreover, in a discernment phase, as shown in drawing 25 , the result of having carried out the face extract of the face of the man in the image based on picture signal S1A supplied from CCD camera 50 in the face extract processing section 90 is supplied to the face recognition processing section 91. In the face

recognition processing section 91, the discriminant function obtained provisionally is tried on the image on various databases, and a face is detected. And what succeeded in detection is outputted as face data. Moreover, it adds to study data by using as non-face data what failed in detection, and study is done further again.

[0193] Hereafter, the GABOA filtering processing in the face extract processing section 90 and the support vector machine in the face recognition processing section 91 are explained to a detail, respectively.

[0194] (4-2-1) It has already turned out that the cell which has selectivity in GABOA filtering processing human being's vision cell to a certain specific bearing exists. This consists of a cell which reacts to a perpendicular line, and a cell reacted to a level line. GABOA filtering is a spatial filter which consists of two or more filters with orientation selectivity like this.

[0195] The space expression of the GABOA filter is carried out by the GABOA function. The GABOA function g (x y) is a degree type [0196].

[Equation 3]

$$g(x, y) = s(x, y)w_r \quad (x, y)$$

[0197] It is alike, and it consists of Carriers s (x y) and the EMBE lobes wr (x y) of

the letter of two-dimensional gauss analysis which consist of a cosine component so that it may be shown.

[0198] Carrier s (x y) is expressed like a bottom type (4) using two or more functions. Here, a coordinate value (u0, v0) expresses spatial frequency, and P expresses the phase of a cosine component.

[0199] It is here and is a degree type [0200].

[Equation 4]

$$s(x, y)=\exp(j(2\pi(u_0x+v_0y)+P)) \quad \cdots\cdots (4)$$

[0201] The carrier boiled and shown is a degree type [0202].

[Equation 5]

$$Re(s(x, y))=\cos(2\pi(u_0x+v_0y)+P)$$

$$Im(s(x, y))=\sin(2\pi(u_0x+v_0y)+P)$$

[0203] It is a real component Re (it is separable into s (x y) and an imaginary component Im (s (x y)).) so that it is alike and may be shown.

[0204] On the other hand, the EMBE lobe which consists of two-dimensional Gaussian distribution is a degree type [0205].

[Equation 6]

$$w_r(x, y)=K\ \exp(-\pi(a^2\ (x-x_0)_r{}^2+b^2($$

[0206] ** -- it is expressed like.

[0207] Here, an axis of coordinates (x0, y0) is the peak of this function, and constants a and b are the scale parameters of Gaussian distribution. Moreover, suffix r is a degree type [0208].

[Equation 7]

$$(x-x_0)_r=(x-x_0)\cos\theta+(y-y_0)\sin\theta$$

$$(y-y_0)_r=-(x-x_0)\sin\theta+(y-y_0)\cos\theta$$

[0209] It is alike and rotation actuation as shown is meant.

[0210] Therefore, a GABOA filter is a degree type [0211] from above-mentioned (4) types and (6) types.

[Equation 8]

$$g(x, y)=K\ \exp(-\pi(a^2(x-x_0)_r{}^2+b^2(y-$$

$$\exp(j(2\pi(u_0x+u_0y)+P))$$

[0186] In addition, in the entry of the 8th line of drawing 23 , one utterance of a non-registered word "Orange" and utterance of a non-registered word "a book" of 19 are clustered by the same cluster. Although it is thought that this cluster should turn into a cluster of a non-registered word "a book" from the utterance used as that member, utterance of a non-registered word "Orange" also serves as a member of that cluster. However, this cluster will also be considered to be clustered by the cluster which cluster division is carried out and makes only utterance of a non-registered word "a book" a member, and the cluster which makes only utterance of a non-registered word "Orange" a member if utterance of a non-registered word "a book" is inputted further after that.

[0187] (4-2) Explain the concrete configuration of the face recognition section 62, next the concrete configuration of the face recognition section 62.

[0188] As shown in drawing 24 and drawing 25 , although the face recognition section 62 can answer in fixed time amount under the environment where it changes dynamically, it consists of the face extract processing section 90 which extracts a face pattern from the inside of the image based on picture signal S1A given from CCD camera 50 (drawing 5), and the face recognition processing section 91 which recognizes a face based on the extracted face pattern. With the gestalt of this operation, "the support vector machine (Support Vector Machine:SVM)" is adopted as face recognition processing in which adopt

[0212] It is alike and is expressed as a space function as shown.

[0213] The direction of eight kinds and three kinds of frequencies are used for the face extract processing section 90 concerning the gestalt of this operation, and it performs face extract processing using a total of 24 GABOA filters.

[0214] Gi is used as the i-th GABOA filter, the result (Gabor Jet) of the i-th GABOA is set to Ji, and an input image is set to I, then the response of a GABOA filter is a degree type [0215].

[Equation 9]

$$J_i(x, y) = G_i(x, y) \oplus I(x, y)$$

[0216] It is come out and expressed. The operation of this (9) type is accelerable using a fast Fourier transform in fact.

[0217] In order to investigate the engine performance of the created GABOA filter, it carries out by reconstructing the pixel filtered and obtained. The reconstructed image H is a degree type [0218].

[Equation 10]

$$H(x, y) = \sum_{i=1}^{0} a_i J_i(x, y)$$

[0219] ** -- it is expressed like.

[0220] And the error E with the input image I and the reconstructed image H is a degree type [0221].

[Equation 11]

$$E = \frac{1}{2} \| I(x, y) - H(x, y) \|^2 = \frac{1}{2} \sum_{x, y} (I(x$$

[0222] ** -- it is expressed like.

[0223] It is reconstructible by asking for optimal a which makes this error E min.

[0224] (4-2-2) Identify that it is the face which corresponds using the support vector machine (SVM) made the highest [study generalization capacity] in the field of pattern recognition about the face recognition in the face recognition processing section 91 with the gestalt of support vector machine book operation.

[0225] About the SVM itself, the report (B. Sholkopf, C.Burges, A.Smola, "Advance in Kernel Support Vector Learning", The MIT Press, 1999.) of the work outside B.sholkopf can be mentioned, for example. The result of the preliminary experiment which the applicant for this patent conducted shows that the face recognition approach by SVM shows a good result compared with the technique

of using principal component analysis (PCA) and a neural network.

[0226] SVM is the learning machine which used the linearity discrimination circuit (perceptron) for the discriminant function, and it can extend to nonlinear space by using a kernel function. Moreover, in study of a discriminant function, it is carried out so that the margin of separation between classes may be taken to max, and since the solution is acquired by solving secondary mathematical programming, it can guarantee theoretically that a global solution can be reached.

[0227] Usually, the problem of pattern recognition is test sample x= (x1, x2--.). It is a degree type [0228] to xn).

[Equation 12]

$$f(x)_j = \sum_{1}^{n} w_j x_j + b$$

[0229] It is asking for discriminant function f (x) come out of and given.

[0230] Here, it is a degree type [0231] about the teacher label for study of SVM.

[Equation 13]

$$y=(y1, y2, \cdots\cdots, yn) \qquad \cdots\cdots (13)$$

[0232] ** -- it sets like.

[0233] Then, it is a degree type [0234] about recognition of the face pattern in SVM.

[Equation 14]

$$y_1 (w^T x_i + b) \geq 1$$

[0235] It can regard as a problem which is boiled and the square of the weight factor w under the shown constraint minimizes.

[0236] The problem which such constraint attached can be solved using Lagrange's undecided constant method. Namely, a degree type [0237]

[Equation 15]

$$L(w, b, a) = \frac{1}{2} \| w \|^2 - \sum_{i=1}^{l} a_i (y_i ((x_i \cdot w) + b) - 1)$$

[0238] It is alike, Lagrange who shows is introduced first, and, subsequently it is a degree type [0239].

[Equation 16]

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial w} = 0 \qquad \cdots\cdots (16)$$

[0240] It is alike, and a partial differential is carried out about each of b and w so that it may be shown.

[0241] Consequently, it is discernment of the face pattern in SVM [0242]

[Equation 17]

$$\max \sum ai - \frac{1}{} \sum aiaiyiyi' xj$$

制約条件：$ai \geqq 0, \sum aiyi = 0$

[0243] It can be alike and can regard as secondary shown plan problems.

[0244] When there are few number of dimensions of a feature space than the number of training samples, the scratch variable xi>=0 is introduced, and it is a degree type [0245] about a constraint.

[Equation 18]

$$y_i (w' x_i + b) \geqq 1 - \xi_i$$

[0246] ** -- it changes like.

[0247] About optimization, it is a degree type [0248].

[Equation 19]

$$\frac{1}{2}||w||^2 + C\sum\xi_i$$

[0249] ********** is minimized.

[0250] In this (19) type, C is a multiplier which specifies how far a constraint is loosened, and needs to determine a value experimentally.

[0251] The problem about the number a of Lagrange is a degree type [0252].

[Equation 20]

$$\max \sum ai - \frac{1}{2} \sum aiaiyiyi' xj$$

$$\text{制約条件} : 0 \leqq ai \leqq C, \sum aiyi = 0$$

[0253] ** -- it is changed like.

[0254] However, a non-line type problem cannot be solved with this (20) type. So, with the gestalt of this operation, the kernel function K (x x3) is introduced, it once maps to the space of high order origin (kernel trick), and linearity

separation is carried out in the space. Therefore, in the original space, it becomes equivalent to having line [non-]-type-dissociated.

[0255] A certain map phi is used for a kernel function, and it is a degree type [0256].

[Equation 21]

$$K(x, y) = \Phi(x)^{T} \Phi(x^{l})$$

[0257] ** -- it is expressed like.

[0258] Moreover, the discriminant function shown in (12) types is also a degree type [0259].

[Equation 22]

$$f(\Phi(x)) = w^{T} \Phi(x) + b$$

$$= \sum a_i y_i K(x, x_i) + b$$

[0260] ** -- it can express like.

[0261] Moreover, it is also related with study and is a degree type [0262].

[Equation 23]

$$\max \sum ai - \frac{1}{2}\sum aiaiyiyixjK(Xi, xj)$$

制約条件：$0 \leqq ai \leqq C, \sum aiyi=0$

[0263] It can be alike and can regard as secondary shown plan problems.

[0264] As a kernel, it is a degree type [0265].

[Equation 24]

$$K(x, x^1)=\exp\left\{-\frac{|x-x^1|^2}{\sigma^2}\right\} \quad \cdots\cdots (24)$$

[0266] It can be alike and the shown Gaussian kernel (RBF (Radius Basic Function)) can be used.

[0267] In addition, you may make it change the class of filter about GABOA filtering according to a recognition task.

[0268] It is redundant to have all the images after filtering as a vector in filtering by low frequency. Then, a down sampling is carried out and you may make it drop the dimension of a vector. 24 kinds of vectors by which the down sampling was carried out are arranged in a single tier, and it is made a long vector.

[0269] Moreover, since it is the discrimination circuit which carries out a feature

space for 2 minutes, SVM applied to recognition of a face pattern in the gestalt of this operation is learned so that it may distinguish "Man A" and "he not being Man A." Therefore, out of the image of a database, Man's A face images are collected first and the label "he is not Man A" is stuck on the vector after GABOA filtering. Generally, more ones of the number of the face images to collect than the dimension of a feature space are good. One discrimination circuit is similarly constituted to each man like "he is Man B" and -- which "is not Man B" to recognize ten persons' face.

[0270] By such study, the support vector which divides "he is not Man A" with "Man A" can be found. SVM is a discrimination circuit which divides a feature space into two, and when a new face image has been inputted, the vector of GABOA filtering outputs a recognition result too by in which of the interface which the support vector for which it asked constitutes it is. And to a boundary, if it is in the field of "Man A", it can be recognized as "Man A." Moreover, if it is the field which "is not Man A", it will be recognized as "He is not Man A."

[0271] The field cut out from the image based on picture signal S1A from CCD camera 50 as a part of a face is not fixed. For this reason, it may be projected on the point which separated with the category to recognize in a feature space. Therefore, a recognition rate may improve by presuming to parts with the descriptions, such as an eye, a nose, and opening, and carrying out morphing by

affine transformation.

[0272] Moreover, the bootstrap technique is employable in order to improve the recognition engine performance. An image is photoed apart from the image used for study, and it uses for a bootstrap. This means feeding the input image into a study set, and relearning it, when the recognition result which the learned discrimination circuit mistook is taken out.

[0273] Moreover, in order to improve the recognition engine performance, there is also a method of seeing time amount change of a recognition result. By the easiest approach, when recognized as 8 times "Man A" among 10 times, it is recognizing it as "Man A" etc. Otherwise, the predicting method which used the Kalman filter is proposed.

[0274] In actuation of the gestalt of this operation, and the configuration beyond effectiveness (5) By this robot 1 While acquiring the man's identifier through a dialogue with a new person, relating the identifier concerned with each data of the acoustical description of the voice of the man who detected based on the output of a microphone 51 or CCD camera 50, and the gestalt-description of a face and memorizing it People's identifier is learned, as an appearance of a new person is recognized to the pan which does not acquire the identifier based on the these-memorized various data, and the new person's identifier, the acoustical description of voice, and the gestalt-description of a face are acquired

like **** and memorized.

[0275] Therefore, without needing the identifier registration by the explicit directions from users, such as an input of a voice command, and press actuation of a touch sensor, this robot 1 can learn the identifier of a new person, a body, etc. automatically through a dialogue with the usual man so that human being may carry out usually.

[0276] According to the above configuration, the man's identifier is acquired through a dialogue with a new person. While relating the identifier concerned with each data of the acoustical description of the voice of the man who detected based on the output of a microphone 51 or CCD camera 50, and the gestalt-description of a face and memorizing it Based on each these-memorized data, recognize an appearance of a new person to the pan which does not acquire the identifier, and acquire the new person's identifier, the acoustical description of voice, and the gestalt-description of a face like ****, and they are memorized. By having learned people's identifier, it can make it possible to learn the identifier of a new person, a body, etc. automatically through a dialogue with the usual man, and the robot which may raise entertainment nature on a target markedly in this way can be realized.

[0277] (6) it is the gestalt of other operations -- although the case where this invention was applied to the robot 1 of the 2-pair-of-shoes walk mold constituted

like drawing 1 in the gestalt of above-mentioned operation was described -- this invention -- not only this -- in addition, in addition to this, it is widely applicable to various equipments other than various robot equipments and robot equipment.

[0278] Moreover, by having a function for conversing with human being in the gestalt of above-mentioned operation, and constituting a dialogue means to acquire the identifier of the target body from human being through the dialogue concerned from the speech recognition section 60, a dialogue control section 63, and the speech synthesis section 64 Although the case where people's identifier was acquired by the voice dialogue with people was described, you may make it this invention constitute a dialogue means so that people's identifier may be acquired by the alphabetic character dialogue not only by this but keyboard entry etc.

[0279] Furthermore, this invention is replaced not only with this but with a person, or, in addition to a person, in addition to this, it may be made to make it into the object of identifier study of the various bodies of an except, although the case where the object of identifier study was a person was described in the gestalt of above-mentioned operation.

[0280] In this case, although the case where recognized that person from the acoustical description of the target person's voice and the gestalt-description of a face, respectively, and it was distinguished in the gestalt of above-mentioned

operation based on these recognition results whether that person is a new person was described Replace this invention not only with this but with this, or, in addition to this, the person is recognized for an individual from two or more kinds of various identifiable descriptions biologically [smells other than these (for example, a form) etc.], respectively. You may make it distinguish whether the person is a new person based on these recognition results. Moreover, when the candidates for identifier study are bodies other than a person, the body is recognized for bodies, such as a color, a configuration, a pattern, and magnitude, from two or more kinds of identifiable descriptions, respectively, and you may make it distinguish whether the body is a new body based on these recognition results. And what is necessary is just to make it establish two or more recognition means to recognize the body made into the object concerned, based on the data of the description with which the detection result concerned and the known body memorized beforehand correspond, while detecting the predetermined description that bodies differ in these cases, respectively.

[0281] Furthermore, in the gestalt of above-mentioned operation, although the case where memory constituted a storage means which associated the identifier of a known body and the recognition result of each recognition means (the speaker-recognition section 61 and face recognition section 62) against the body concerned to relate and to memorize information was described This invention

can apply widely various storage means other than the memory which can memorize not only this but information (for example, a disk-like record medium etc.) in addition to this.

[0282] Furthermore, in the gestalt of above-mentioned operation, although the case where the recognition processing the speaker-recognition section 61 and the face recognition section 62 recognize the target man to be was made not to be performed only once was described This invention may be made to be made to perform recognition processing once again, not only this but when it is for example, recognition impossible (SID=-1), and it may be it at the times other than this, or it may be made to perform recognition processing of multiple times. The precision of a recognition result can be raised by doing in this way.

[0283] Furthermore, in the gestalt of above-mentioned operation, by the majority of the recognition result of the recognition means (the speech recognition section 60, the speaker-recognition section 61, face recognition section 62) of plurality [control section / 63 / dialogue], although the case where it was judged whether you are a person with the new man was described The man may be made, as for this invention, to judge whether you are a new person based on each recognition result of the recognition means of these plurality by technique not only this but other than majority.

[0284] In this case, set, for example, weighting is made each recognition result

of two or more recognition means according to the recognition engine performance of that recognition means. The approach of judging whether the target body being new based on each of that recognition result that carried out weighting, When it is able to be most judged as a new person based on the recognition result of the high recognition means of the recognition engine performance, and other one recognition means, various approaches, such as ****** which does not use the recognition result of other recognition means, can be applied widely.

[0285] Furthermore, in the gestalt of above-mentioned operation, although the case where it was made to raise recognition precision according to statistical stability by [the] reaching speaker-recognition section 61 or making the face recognition section 62 carry out additional study was described when the speaker-recognition section 61 and the face recognition section 62 have recognized the target man correctly You may make it this invention prepare the function which may raise the dependability of the correlation information by memorizing the same combination repeatedly similarly about the correlation information stored not only in this but in the memory 65. Specifically, the approach using the neural network indicated by "the Institute of Electronics, Information and Communication Engineers paper magazine, D-II, Vol.J82-D-II, No6, pp.1072-1081." can be used as the embodiment approach of such a

function.

[0286]

[Effect of the Invention] A dialogue means to acquire the identifier of the target body through a dialogue in study equipment as mentioned above according to this invention, Two or more recognition means to recognize the target body based on the data of the description with which the data of two or more descriptions of the target body are detected, and the detection result concerned and a known body correspond, A storage means which associated the recognition result of each recognition means against the identifier of a known body to relate and to memorize information, A decision means to judge whether the target body is a new body based on the identifier of the body made into the object which the dialogue means acquired, the recognition result of each recognition means against the target body, and the correlation information that a storage means memorizes, When a decision means judges the target body to be a new body, while making each recognition means memorize the data of two or more descriptions corresponding to the body made into the object concerned By having established the control means which it relates [control means] about the body made into the object concerned, and makes a storage means memorize information The identifier of a new person, a body, etc. can be automatically learned through a dialogue with the usual man, and the study equipment which

may raise entertainment nature on a target markedly in this way can be realized so that human being may carry out usually.

[0287] Moreover, the dialogue step which acquires the identifier of the target body through a dialogue in the study approach according to this invention, Two or more recognition steps which recognize the target body based on the data of the description with which the data of two or more descriptions of the target body are detected, and the detection result concerned and a known body correspond, The storage step which associated the recognition result of each recognition means against the identifier of a known body and which relates and memorizes information, The decision step which judges whether the target body is a new body based on the identifier of the body made into the object which the dialogue means acquired, the recognition result of each recognition means against the target body, and the correlation information that a storage means memorizes, When a decision means judges the target body to be a new body, while making each recognition means memorize the data of two or more descriptions corresponding to the body made into the object concerned By having prepared the control step which it relates [step] about the body made into the object concerned, and makes a storage means memorize information The identifier of a new person, a body, etc. can be automatically learned through a dialogue with the usual man, and the study approach which may raise entertainment nature on

a target markedly in this way can be realized so that human being may carry out usually.

[0288] A dialogue means to acquire the identifier of the target body through a dialogue in robot equipment furthermore according to this invention, Two or more recognition means to recognize the target body based on the data of the description with which the data of two or more descriptions of the target body are detected, and the detection result concerned and a known body correspond, A storage means which associated the recognition result of each recognition means against the identifier of a known body to relate and to memorize information, A decision means to judge whether the target body is a new body based on the identifier of the body made into the object which the dialogue means acquired, the recognition result of each recognition means against the target body, and the correlation information that a storage means memorizes, When a decision means judges the target body to be a new body, while making each recognition means memorize the data of two or more descriptions corresponding to the body made into the object concerned By having established the control means which it relates [control means] about the body made into the object concerned, and makes a storage means memorize information The identifier of a new person, a body, etc. can be automatically learned through a dialogue with the usual man, and the robot equipment which

may raise entertainment nature on a target markedly in this way can be realized so that human being may carry out usually.

[0289]

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is the perspective view showing the appearance configuration of the robot by the gestalt of this operation.

[Drawing 2] It is the perspective view showing the appearance configuration of the robot by the gestalt of this operation.

[Drawing 3] It is the approximate line Fig. with which explanation of the appearance configuration of the robot by the gestalt of this operation is presented.

[Drawing 4] It is the approximate line Fig. with which explanation of the internal configuration of the robot by the gestalt of this operation is presented.

[Drawing 5] It is the approximate line Fig. with which explanation of the internal configuration of the robot by the gestalt of this operation is presented.

[Drawing 6] It is the block diagram with which explanation of processing of the

Maine control section 40 about an identifier learning function is presented.

[Drawing 7] It is the conceptual diagram with which explanation of correlation with FID and SID, and the identifier in memory is presented.

[Drawing 8] It is the flow chart which shows identifier study procedure.

[Drawing 9] It is the flow chart which shows identifier study procedure.

[Drawing 10] It is the approximate line Fig. showing the example of a dialogue at the time of identifier study processing.

[Drawing 11] It is the approximate line Fig. showing the example of a dialogue at the time of identifier study processing.

[Drawing 12] It is the conceptual diagram with which the explanation of new registration with FID and SID, and an identifier is presented.

[Drawing 13] It is the approximate line Fig. showing the example of a dialogue at the time of identifier study.

[Drawing 14] It is the approximate line Fig. showing the example of a dialogue at the time of identifier study processing.

[Drawing 15] It is the block diagram showing the configuration of the speech recognition section.

[Drawing 16] It is the conceptual diagram with which explanation of a word dictionary is presented.

[Drawing 17] It is the conceptual diagram with which explanation of the syntax

rule is presented.

[Drawing 18] It is the conceptual diagram with which explanation of the contents of storage of a feature-vector buffer is presented.

[Drawing 19] It is the conceptual diagram with which explanation of a score sheet is presented.

[Drawing 20] It is the flow chart which shows speech recognition procedure.

[Drawing 21] It is the flow chart which shows a non-registered word-processing procedure.

[Drawing 22] It is the flow chart which shows cluster division procedure.

[Drawing 23] It is the conceptual diagram showing a simulation result.

[Drawing 24] It is the block diagram showing the configuration of the face recognition section at the time of study.

[Drawing 25] It is the block diagram showing the configuration of the face recognition section at the time of recognition.

[Description of Notations]

1 [.. A microphone, 54 / .. A loudspeaker, 60 / .. The speech recognition section, 61 / .. The speaker-recognition section, 62 / .. The face recognition section, 63 / .. A dialogue control section, 64 / .. The speech synthesis section, 65 / .. Memory, S1A / .. A picture signal, S1B, S3 / .. A sound signal, D1, D2 / .. Character-string data, RT1 / .. Identifier study procedure.] .... A robot, 40 .. The Maine control

section, 50 .. A CCD camera, 51